AI for Medicine and Health @ Bozen-Bolzano

Giuseppe Di Fatta, Andrea Janes, Paola Lecca,
Fabrizio Maria Maggi, Marco Montali, Floriano Zini
Faculty of Computer Science, Free University of Bozen-Bolzano, Italy
{giuseppe.difatta, andrea.janes, paola.lecca, floriano.zini}@unibz.it
{maggi, montali}@inf.unibz.it

Abstract

This document outlines the research carried out at the Faculty of Computer Science of the Free University of Bozen-Bolzano that applies artificial intelligence to the field of medicine and health. A substantial part of the activities is devoted to the use of machine learning techniques: in the prediction of neurodegenerative diseases such as Alzheimer's; in the recognition of sports activities from data collected by sensors to support training and rehabilitation; in making analysis of genetic data; and in producing personalization and adaptation in computerized cognitive training paths. Moreover, our research investigates how to profitably exploit hybrid process modelling in order to monitor the adherence to clinical guidelines. For each research area, we also report the collaborations with other institutions, and the related ongoing or recently completed projects.

1 Machine Learning for the Prediction of Neurodegenerative Diseases

Involved people

Giuseppe Di Fatta (Faculty of Computer Science, Free University of Bozen-Bolzano, Italy)
Ali Varzandian, Miguel Angel Sanchez Razo, Akhila Atmakuru (University of Reading, UK)

Projects

The ongoing research activities have been supported by the University of Reading (UK) with a PhD scholarship (€100,000) and by KNIME (CH) with an industry-funded research grant (€25,000).

Description

Alzheimer's disease (AD) is a terminal neurodegenerative disease and the most common type of dementia. Currently, there is no cure and no single diagnostic test for AD. Moreover, it is believed that the pathophysiological process starts many years before any evident cognitive decline. The potential benefits of an early diagnosis are matched by its difficulty due to a long asymptomatic stage and the lack of definitive biomarkers. Machine learning algorithms ([Sarica *et al.*, 2014b; Sarica *et al.*, 2014a; Bron and others, 2015; Spedding *et al.*, 2015a; Spedding *et al.*, 2015b]) can be applied to brain images for the diagnosis of AD and other neurodegenerative diseases. Brain structural magnetic resonance imaging (MRI) can unveil atrophy in different regions of the brain. In particular, imaging biomarkers ([Varzandian *et al.*, 2021]) based on from brain-imaging technologies can provides the opportunity to generate accurate diagnostic tools as well as useful insights, thanks also to the availability of large multi-source data. This research activity intends to identify and evaluate suitable machine learning approaches for such type of biomarkers to achieve three important objectives: high predictive accuracy, useful explainability, and specificity with respect to various neurodegenerative conditions.

2 Activity Recognition in Sports to Support Training, Health, and Rehabilitation

Involved people

Andrea Janes, Marina Andric, Sandeep Gupta, Iustina Ivanova, Sadaf Moaveninejad, Francesco Ricci, Ahmet Onur Yolcu, Floriano Zini (Faculty of Computer Science, Free University of Bozen-Bolzano, Italy)

Projects

This research has been supported through two projects: "Internet of Things for Climbers (IoTforC), an industry funded project (5,000 €), and "Sensors and data for the analysis of sports activities (SALSA)", funded by the European Regional Development Fund (ERDF) (465,000 €).

Description

Recording and measuring athletic performance through sensors installed either on the athlete's body (e.g., in the form of a bracelet) or on sports equipment is a method for tracking fitness goals. Sensors must record the athlete's movements in such a way that distinguished events can be detected in the subsequent data analysis, enabling statements to be made about the success of the training. Our research aims at developing different prototypes (consisting of a sensor component and a software component) which can be used in sports to record and evaluate athletes' movements. The tracked movements can be than used to determine performance, monitor

a rehabilitation process, understand training progress, recommend routes to achieve training goals and so on.

In the SALSA project the climbing sport is our primary application in which the different technologies can be tested [Ivanova *et al.*, 2020a; Ivanova *et al.*, 2020b; Ivanova *et al.*, 2022]. In SALSA we adopt state-of-the art supervised and unsupervised Machine Learning algorithms to perform the activity recognition task. The prototypes obtained vary in their application (outdoor, indoor, on the athlete as a "wearable", i.e. as a portable sensor), as well as in the possibility of data analysis. Each prototype represents a model solution that can be used in future projects/products to collect and process data on the activities performed by athletes.

3 Altered Gene Pathways in Chronic Myeloid Leukemia Patients

Involved people

Paola Lecca, Giulia Lombardi (Faculty of Computer Science, Free University of Bozen-Bolzano, Italy)
Davide Ferrari, Zhendong Huang (Faculty of Economics and Management, Free University of Bozen-Bolzano, Italy)
Claudio Sorio (Department of Medicine, University of Verona, Italy)

Projects

The Faculty of Computer Science funded two ongoing projects: COMPANET (Comparative analysis for detecting altered pathways in gene networks), and DAQETA-CML (Detecting and quantifying (side-)effects of recent experimental therapies against Chronic Myeloid Leukemia), supported by a budget of 10,000 €, and 82,000 €, respectively.

Description

The main purpose of COMPANET is the detection through computational procedures of pathways in gene networks that are most likely affected by the state of the gene coding for Protein Tyrosine Phosphatase, Receptor Type, G (PTPRG), which is a signalling molecule regulating cell growth, differentiation, mitotic cycle, and oncogenic transformation. The study considers also the effect of Tyrosine Kinases Inhibitors (TKI) both in presence and absence of activity of PTPRG. Inhibition of tyrosine kinases represents a strategy to disrupt signalling pathways that promote neoplastic growth and survival in hematologic malignancies, and likely in other neoplasias as well. The study uses gene expression data in healthy patients and in patients with chronic myelogenous leukemia under different conditions related to the status of PTPRG (active/inactive) and the presence of TKI. The conditions are expected to be reflected in networks with different topologies, different edge weights and different expression levels of the genes, that in turn reflect in different distributions of the well-known centrality measures, such as betweenness, clustering coefficient, and new emerging ones such as vibrational centrality and expected force, which recently have gained attention due to their ability to quantify node robustness and spreading information, respectively. The main results of the project are new graph analytics algorithms for the analysis of the static and dynamic properties of the network topology under different conditions and new approaches to differential expression analysis [Lecca and Lombardi, 2021], as well as new algorithms based on AI techniques for causal inference [Lecca, 2021] in cancer biology and model calibration.

The DAQETA-CML project concerns the analysis of the genome of patients affected by Chronic Myeloid Leukemia (CML) with the two-fold aim of identifying the gene features and the gene interactions associated with the onset of the disease and detecting genomic alterations due to the effect of experimental therapeutic treatments. This project integrates two contributions: new context-specific biological network analysis techniques, and new statistical hypothesis testing frameworks adequate to process incomplete data and data that do not satisfy the assumption of commonly used hypothesis tests. The tests are used to compare the topological properties of gene networks that underwent different pharmacological treatments. The final objective is to identify genes and pathways affected by new experimental drugs.

4 Hybrid Processes for Modeling and Monitoring Clinical Guidelines

Involved people

Fabrizio Maria Maggi, Marco Montali, Andrey Rivkin (Faculty of Computer Science, Free University of Bozen-Bolzano, Italy)
Anti Alman (University of Tartu, Estonia), Fabio Patrizi (Sapienza University of Rome)

Description

A key functionality of any process-aware information system is *monitoring* [Ly *et al.*, 2015]. Monitoring concerns the ability to detect, and therefore handle, deviations appearing in ongoing process instances and, in most cases, requires the expected behavior of the process to be specified in advance. Such specifications are commonly created in the form of procedural or declarative process models, depending on the scenario at hand. In general, procedural models are more suitable for relatively structured processes (e.g., an automated manufacturing line), while declarative models are more suitable for flexible and knowledge-intensive processes (e.g., the management of a natural disaster).

The majority of existing monitoring approaches rely on the assumption that the full knowledge-base (control-flow, decision rules, temporal aspects, etc.) required for monitoring each possible process instance can be embedded into a single process specification. While this is a reasonable assumption for processes with homogeneous behavior (i.e., either structured or knowledge-intensive), there are domains in which processes are characterized by a combination of several (independently defined) procedural sub-processes working under additional (declarative) domain constraints. A prime example of this can be found in the medical domain as witnessed by state-of the art approaches like [Spiotta *et al.*, 2017; Bottrighi *et al.*, 2011]. Here, patients with co-morbidities may be subject to multiple clinical pathways at once (consisting of various medical actions) that can be enriched with additional, context-dependent constraints. Monitoring such

multi-model scenarios becomes a challenging task as it requires not only to account for checking a process instance against multiple heterogeneous models, but also to provide mechanisms for handling the interplay of those models.

For example, prescribing a certain treatment without taking into consideration patient's preexisting conditions may result in adverse effects causing a worsening of the patient's state. Ideally, these conflicts should be detected at early stages of the process execution and reported to health providers so as to avoid undesired treatment outcomes.

To address these problems, we use a Multi-Model Monitoring Framework for hybrid process specifications (M3 Framework) presented in [Alman *et al.*, 2021]. The framework consists of multiple phases, starting from knowledge elicitation prior to process execution, and ending with the successful completion of a scenario specific monitoring task. The monitoring task can seamlessly handle the interplay of process specifications, detect inevitable violations in advance, and provide recommendations on the next course of action based on a violation cost model (either avoiding violations or, if that is not possible, minimizing the total cost of violations).

The key benefits of the M3 Framework are (1) support for monitoring process executions with respect to multiple process specifications simultaneously, thus circumventing the need to embed the full knowledge-base required for the execution of every possible process instance into a single process specification, and (2) support for both procedural and declarative process specifications, thus circumventing the need to force declarative knowledge into procedural specifications (and vice versa).

5 Adaptive Cognitive Training

Involved people

Floriano Zini (Faculty of Computer Science, Free University of Bozen-Bolzano, Italy)
Fabio Le Piane, Mauro Gaspari (Department of Computer Science and Engineering, University of Bologna, Italy)

Description

Computer-assisted cognitive training can help patients affected by several illnesses alleviate their cognitive deficits, or healthy people improve their mental performance. Existing systems for computerized cognitive training generally embed mechanisms for the execution of graded exercises, whose difficulty increases gradually according to a set of predefined rules, considering the results the trainees achieve, that are the same for all individuals. Even though these systems are often defined as adaptive, they do not take into account the specificities of the trainees. Our research is contributing to computerized cognitive training systems that incorporate adaptive mechanisms for cognitive training that can instead fully personalize the training path to the single individual. We have realized MS-rehab [Gaspari *et al.*, 2020; MSR, 2022], a computerized cognitive rehabilitation system dedicated to patients suffering from multiple sclerosis, but usable also by subjects affected by other pathology or by healthy people. MS-rehab can be used to implement and monitor the rehabilitation process both at the hospital and at home. The system standardizes the training process and monitor the patient's activity by recording the parameters of execution, the outcome of the performance and the progress of the rehabilitation process over time. We have used Reinforcement Learning (RL) to learn how to automatically adapt the difficulty of computerized training exercises proposed by MS-rehab to the trainee [Zini *et al.*, 2021]. Our approach is to associate a personalized RL agent to each trainee, for each cognitive exercise she/he performs. The agent interacts with the subject while she/he is executing the exercise, and learns from this experience a personalized policy able to vary the exercise difficulty according to how the trainee is performing. We propose a two-phase method for eliciting difficulty variation policies to be embedded in a cognitive training system. The purpose of the first phase is to build an exercise policy that is specific to a category of trainees. The second phase is dedicated to refining this policy and tailoring it to an individual in the category. We have realized two user studies whose results provide evidence of the effectiveness of our method: a first study, in which a student category policy obtained via RL was found to have better effects on the cognitive function than a standard baseline training that adopts a mechanism to vary the difficulty proposed by neuropsychologists; and a second study, demonstrating that adding an RL-based individual customization further improves the training process. We are currently designing other experiments involving patients suffering from Attention Deficit Hyperactivity Disorder, in which we want to further evaluate our RL approach and test its robustness.

References

[Alman *et al.*, 2021] Anti Alman, Fabrizio Maria Maggi, Marco Montali, Fabio Patrizi, and Andrey Rivkin. Monitoring hybrid process specifications with conflict management: The automata-theoretic approach. *CoRR*, abs/2111.13136, 2021.

[Bottrighi *et al.*, 2011] Alessio Bottrighi, Federico Chesani, Paola Mello, Marco Montali, Stefania Montani, and Paolo Terenziani. Conformance checking of executed clinical guidelines in presence of basic medical knowledge. In Florian Daniel, Kamel Barkaoui, and Schahram Dustdar, editors, *Proceedings of the 2011 Business Process Management Workshops*, volume 100 of *Lecture Notes in Business Information Processing*, pages 200–211. Springer, 2011.

[Bron and others, 2015] Esther E. Bron et al. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural mri: The caddementia challenge. *NeuroImage*, 111:562–579, 2015.

[Gaspari *et al.*, 2020] Mauro Gaspari, Floriano Zini, and Sergio Stecchi. Enhancing cognitive rehabilitation in multiple sclerosis with a disease-specific tool. *Disability and Rehabilitation: Assistive Technology*, 0(0):1–14, 2020. PMID: 33259243.

[Ivanova *et al.*, 2020a] Iustina Ivanova, Marina Andric, Andrea Janes, Francesco Ricci, and Floriano Zini. Climbing activity recognition and measurement with sensor data

analysis. In *Companion Publication of the 2020 International Conference on Multimodal Interaction*, ICMI '20 Companion, page 245–249, New York, NY, USA, 2020. Association for Computing Machinery.

[Ivanova *et al.*, 2020b] Iustina Ivanova, Marina Andrić, Sadaf Moaveninejad, Andrea Janes, and Francesco Ricci. Video and sensor-based rope pulling detection in sport climbing. In *Proceedings of the 3rd International Workshop on Multimedia Content Analysis in Sports*, MMSports '20, page 53–60, New York, NY, USA, 2020. Association for Computing Machinery.

[Ivanova *et al.*, 2022] Iustina Ivanova, Marina Andrić, and Francesco Ricci. Content-based recommendations for crags and climbing routes. In Jason L. Stienmetz, Berta Ferrer-Rosell, and David Massimo, editors, *Information and Communication Technologies in Tourism 2022*, pages 369–381, Cham, 2022. Springer International Publishing.

[Lecca and Lombardi, 2021] Paola Lecca and Giulia Lombardi. Identification of genes responsive to treatments against chronic myeloid leukemia. https://gitlab.inf.unibz.it/Paola.Lecca/chronic-myeloid-leukemia-genes, 2021.

[Lecca, 2021] Paola Lecca. Machine learning for causal inference in biological networks: Perspectives of this challenge. *Frontiers in Bioinformatics*, 1, September 2021.

[Ly *et al.*, 2015] Linh Thao Ly, Fabrizio Maria Maggi, Marco Montali, Stefanie Rinderle-Ma, and Wil M. P. van der Aalst. Compliance monitoring in business processes: Functionalities, application, and tool-support. *Inf. Syst.*, 54:209–234, 2015.

[MSR, 2022] Ms-rehab - riabilitazione cognitiva computerizzata nella sclerosi multipla. https://rehab.cs.unibo.it/MS-rehab-website/, 2022.

[Sarica *et al.*, 2014a] A. Sarica, G. Di Fatta, and M. Cannataro. K-Surfer: A KNIME extension for the management and analysis of human brain mri FreeSurfer/FSL data. In *Brain Informatics and Health (BIH), 2014. Lecture Notes in Computer Science, vol 8609. Springer*, 2014.

[Sarica *et al.*, 2014b] A. Sarica, G. Di Fatta, G. Smith, M. Cannataro, and J. D. Saddy. Advanced feature selection in multinominal dementia classification from structural mri data. *Proc MICCAI Workshop Challenge on Computer-Aided Diagnosis of Dementia Based on Structural MRI Data*, pages 82–91, 2014.

[Spedding *et al.*, 2015a] Alexander Luke Spedding, Giuseppe Di Fatta, and Mario Cannataro. A genetic algorithm for the selection of structural mri features for classification of mild cognitive impairment and alzheimer's disease. In *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1566–1571, 2015.

[Spedding *et al.*, 2015b] Alexander Luke Spedding, Giuseppe Di Fatta, and James Douglas Saddy. An LDA and probability-based classifier for the diagnosis of alzheimer's disease from structural mri. In *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1404–1411, 2015.

[Spiotta *et al.*, 2017] Matteo Spiotta, Paolo Terenziani, and Daniele Theseider Dupré. Temporal conformance analysis and explanation of clinical guidelines execution: An answer set programming approach. *IEEE Trans. Knowl. Data Eng.*, 29(11):2567–2580, 2017.

[Varzandian *et al.*, 2021] Ali Varzandian, Miguel Angel Sanchez Razo, Michael Richard Sanders, Akhila Atmakuru, and Giuseppe Di Fatta. Classification-biased apparent brain age for the prediction of alzheimer's disease. *Frontiers in Neuroscience*, 15, 2021.

[Zini *et al.*, 2021] Floriano Zini, Fabio Le Piane, and Mauro Gaspari. Adaptive cognitive training with reinforcement learning. *ACM Transactions on Interactive Intelligent Systems*, 2021. Accepted for publication.