Machine learning and deep analytics for biocomputing: call for better explainability

Dragutin Petkovic[*] [†]
Computer Science Department, San Francisco State University (SFSU)
1600 Holloway Ave, San Francisco, CA 94132
Petkovic@sfsu.edu

Lester Kobzik
Department of Environmental Health, Harvard University
665 Huntington Avenue, Boston, Massachusetts 02115
lkobzik@hsph.harvard.edu

Christopher Re
Department of Computer Science, Stanford University
353 Serra Mall, Stanford, CA 94305
chrismre@cs.stanford.edu

The goals of this workshop are to discuss challenges in *explainability* of current Machine Leaning and Deep Analytics (MLDA) used in biocomputing and to start the discussion on ways to improve it. We define explainability in MLDA as *easy to use* information explaining *why and how the MLDA approach made its decisions*. We believe that much greater effort is needed to address the issue of MLDA explainability because of: 1) the ever increasing use and dependence on MLDA in biocomputing including the need for increased adoption by non-MLD experts; 2) the diversity, complexity and scale of biocomputing data and MLDA algorithms; 3) the emerging importance of MLDA-based decisions in patient care, in daily research, as well as in the development of new costly medical procedures and drugs. This workshop aims to: a) analyze and challenge the current level of explainability of MLDA methods and practices in biocomputing; b) explore benefits of improvements in this area; and c) provide useful and practical guidance to the biocomputing community on how to address these challenges and how to develop improvements. The workshop format is designed to encourage a lively discussion with panelists to first motivate and understand the problem and then to define next steps and solutions needed to improve MLDA explainability.

Keywords: Machine Learning, explainability, interpretability, workshop

1. Introduction, Background and Motivation

The goals of this workshop are to discuss challenges in *explainability* of current Machine Leaning and Deep Analytics (MLDA) used in biocomputing and to explore ways to improve it.

*Travel partially supported by Mobilize Center, Stanford University

We define explainability in MLDA as *easy to use* information explaining *why and how the MLDA approach made its decision.* Successful explainability will offer much deeper insights into MLDA operation compared to what is available today. Algorithms and software implementing MLDA decision models are inherently complex and notoriously difficult to understand and communicate. This creates barriers to their adoption by non-experts and challenges in their validation, reproducibility and benchmarking in the research community. The input data ("training databases") have a critical influence on MLDA results but are complex, and they change with time as more and more or better measurements and samples are added. Frequently, a "gold standard" or ground truth is not easily available. All this makes it very challenging to understand, evaluate, verify and even reproduce results of published MLDA work. At the same time, a review of the literature shows that very few research efforts and methods focus specifically on MLDA explainability.

Interest in explaining how ML systems work is growing within not only funding agencies and potential adopters but also in general public reflecting the penetration of ML in all aspects of our lives. Specifically, we believe that improved explainability of MLDA in biocomputing will result in the following benefits: increased credibility and confidence in its application; improved ability to objectively evaluate, audit and verify MLDA solutions; and possible discovery of new knowledge and ideas enabled by better understanding of how MLDA works on specific problems.

2. Workshop Format and Organization

Six workshop panelists will represent all four constituencies in the biocomputing ecosystem: 1) *computational researchers who are experts in MLDA* and who develop and use the technology; 2) *biocomputing practitioners* who are using MLDA but are not experts; 3) *editors/ evaluators* who need to decide what to publish; 4) and members of *funding agencies* who evaluate research results and use the funding to influence the direction of research. The 3-hour workshop is organized in the form of two main panels, followed by a discussion. The panelists are:

- A. Esteva *(Ph. D. Candidate, Stanford University)*
- Dr. R. Ghanadan (*Google since September 2017; formerly Program Manager, Defense Sciences Office, DARPA*)
- Dr. W. Kibbe (*Chief for Translational Biomedical Informatics in the Department of Biostatistics and Bioinformatics and chief data officer for the Duke Cancer Institute, Professor, Duke University since August 2017; formerly Dir. of NCI Center for Biomedical Informatics and Inf. Technology since*)
- Dr. B. Percha *(Assistant Professor, Icahn School of Medicine at Mount Sinai; Head of R&D, Health Data and Design Innovation Center (HD2i) Institute for Next-Generation Healthcare)*
- Dr. R. Roettger *(Assistant Professor, University of Southern Denmark, Odense)*
- Dr. R. Scheuermann (*Dir. Of Bioinformatics, J. Craig Venter Institute*),

Panel 1: Needs for Explainability in ML and Deep Analytics - View of "Users "
Goal of this section is that panelists who are users but not necessarily experts or developers of MLDA: a) outline their experience, needs and motivation for better explainability in MLDA; and b) encourage and *challenge* developers of MLDA technology to provide better explainability.
Panel moderator: Dr. Lester Kobzik
Panelists: Dr. R. Ganadhan , Dr. B. Percha, Dr. W. Kibbe

Panel 2: Toward Better Explainability in ML and Deep Analytics – View of "Developers"
The goal of this section is for panelists on the development and research side of MLDA techniques to: a) present examples of the state-of-the-art in MLDA explainability; and b) discuss ways to address the challenges outlined by the previous panelists.
Panel moderator: Dr. Christopher Re
Panelists: A. Esteva, Dr. R. Roetger, Dr. R. Scheuermann

Discussion with panelists and audience
Moderator: Dr. D. Petkovic

3. Panelists' Abstracts

In this section we list panelist' abstracts reelecting some of their initial thoughts and ideas to be discussed at the workshop.

AI in healthcare: a case study in explainability

Andre Esteva, Stanford University

In a recent paper we demonstrate classification of skin lesions using a single deep convolutional neural networks (CNN), trained end-to-end from images directly, using only pixels and disease labels as inputs. We train a CNN using a dataset of 129,450 clinical images—two orders of magnitude larger than previous datasets — consisting of 2,032 different diseases. We test its performance against 21 board-certified dermatologists on biopsy-proven clinical images with two critical binary classification use cases: malignant carcinomas versus benign seborrheic keratoses; and malignant melanomas versus benign nevi. An algorithm known as t-SNE is effective at visualization high-dimensional data - we employ it to understand how the algorithm clusters images into disease categories based on visual and clinical similarity. Additionally, we render saliency maps of several example images in order to demonstrate the individual pixels that most influence a trained model's prediction - this is done by backpropagating the gradient to the input layer. Finally, we calculate confusion matrices for the CNN and two board-certified dermatologists on our validation set categories, demonstrating that the CNN misclassifies lesions in a manner similar to experts.

Machine Learning to Machine Understanding, the Need for Explainable AI

Dr. Reza Ghanadan, Google

Machine learning has shown dramatic success across many Artificial Intelligence application areas in recent years, leveraging advances in computing power and the availability of large sets of training data. As an engine for the 4th industrial revolution, AI provides a tremendous opportunity to deploy autonomous systems in many complex and interactive tasks, such as personalized medicine and healthcare, to analyze, learn, decide, and act in complex situations. However, it is essential for the users to be able to understand and trust the decisions of emerging generation of artificial intelligence systems. Productization and wide acceptance of current systems are limited because of our inability to validate and verify their performance when they act in new situations, due in turn to machine's current inability to explain its decisions and actions to human users. To gain our trust, machine-learning systems will need to have the ability to explain their rationale in meaningful ways, characterize trade-offs, and convey an understanding of how they will behave in the future in new situations. Such intrinsic capability would help users and developer community with increasingly more powerful tools and applications. In this talk, we will describe a user's perspective and needs for such capability in emerging ML/AI systems, and highlight a few examples and tools for healthcare and biocomputing at Google.

Machine Learning, Deep Analytics, and support for a Learning Health System

Dr. Warren A. Kibbe, Ph.D, Duke University

There has been a lot of progress in applying MLDA techniques, especially in the field of imaging. Image analysis using MLDA approaches are hitting mainstream computing, as evidenced by the ability of the Photos app in the Mac OS High Sierra to classify pictures containing a 'beach', or 'trees' or 'flowers' and of course face recognition. In general, MLDA techniques have two phases – a feature identification phase, where features are extracted/identified, and an associative learning component, where various statistical techniques are used to identify features that correlate with attributes available in the training set. Using techniques that are 'explainable' for associative learning is highly desired in healthcare, especially when applying these techniques to complex biological data such as whole exome sequencing and RNAseq. Coupling MLDA with Natural Language Processing for classification and decision support processes will increase the value of data in healthcare. For these techniques to reach their potential, explainability is a key need.

Explainability Tales from the Health Entrepreneur Partners Program

Dr. Bethany Percha, Icahn School of Medicine at Mount Sinai; HD2i

At HD2i, we believe many of the technologies that will fuel next-generation healthcare will come from outside the traditional healthcare ecosystem. Often with the right application of machine learning and data science, products geared toward consumers or other industries, such as fitness, can be reoriented to address powerful clinical goals. We engage in data science partnerships with early-stage companies to help bring these fresh ideas and products into the clinic faster. One result of this is that we have had to think hard about how best to explain technical concepts from machine learning and statistics to folks with little previous exposure. In my talk, I'll share some stories from our first partnerships and discuss how what we've learned

through our interactions with industry can help inform broader concepts of explainability in machine learning.

On the Explainability of Clustering Results

Dr. Richard Roettger, University of Southern Denmark, Odense

In the recent years we have seen a tremendous growth in the amount, the complexity, and the diversity of biological data. Often, the first line of defense when facing this amount of data is clustering. But how reliable are clustering results when it is unclear whether the inherit model of the clustering tool fits the data? With ClustEval, we have automatized most steps of a cluster analysis which allowed us to provide a better overview of the existing clustering tools and their respective performances and we could demonstrate how sensitive and erratic some algorithms behave under certain conditions. Furthermore, we present the tool TiCoNE (time course network enrichment) which interactively involves the user in the machine learning process to create a time series clustering. In order to gain explainability and assess the validity of the clustering, the results are enriched with biological networks in order to extract connected biological components behaving consistent over time for a certain condition. This talk should serve as one example of gaining information and explanatory power by means of the integration of independent evidence instead of creating ever more complicated computational models. All resources are available at https://clusteval.compbio.sdu.dk and https://ticone.compbio.sdu.dk

Use of machine learning-derived gene expression features to explain the unique characteristics of cell types defined using single cell RNA sequencing

Dr. Richard H. Scheuermann, J. Craig Venter Institute

Machine learning has become an important instrument in the bioinformatics tool kit, with many different applications associated with various omics technologies. In our single cell transcriptomics program, we use machine learning for sample classification to identify poor quality samples and to help partition cell types using gene expression data from complex cell mixtures. In addition to these classification objectives, we are also finding that methods, like Random Forest, that provide quantitative information about features that are most useful for classification are equally useful for explaining important relationships between features and classes. In the case of sample quality classification, features that are useful in identifying poor quality samples also point to potential problematic steps in the experimental workflow that can then be targeted for process improvement. In the case of cell type classification, gene expression features that are useful for cell type partitioning are proving informative for identifying the necessary and sufficient characteristics for defining discrete cell types, and for illuminating the important biological distinctions that characterize unique cellular phenotypes. These examples highlight the value of capturing explainability information during the machine learning process. This work is supported by the Allen Institute for Brain Science, the JCVI Innovation Fund, and the U.S. National Institutes of Health 1R21AI122100.