FONDAZIONE BRUNO KESSLER
MAGAZINE

What if data is all we have?

March 2, 2017

Studying machine learning and deep learning techniques to develop a novel and original approach for the integration of phenotypic and genotypic data in Autism Spectrum Disorder.

When I use words such as prediction, data analysis, statistically significant, and others of the same sort, my fellow math collegues usually give me the “dirty look”. For a mathematician, applied statistics and data mining cannot be considered as logic and rational as other subjects, such as algebra, geometry, analysis. If your work consists in discovering patterns in large datasets, you are not a mathematician, you are a magician. You might as well pull a rabbit from a hat.

Uncertanty and lack of rigorous proofs are scary. As a mathematician I share the same feelings on the subject. Mathematicians tend to be prejuditial towards the application of statistics. But then I wonder: “What if data is all we have?”. Nature is an imperfect structure and human beings, as a manifestation of it, are imperfect as well. Hence, it is difficult for a system as rigorous as mathematics to be successfully applied to natural and human phenomena. However, some events are understandable and may be modeled by mathematical rules (e.g. the spread of an epidemics). Moreover, information collected in the past, may be useful to predict what will happen in the future (everybody knows that weather forecasts are [not always] magic predictions). Hence, applied mathematics gives the possibility to investigate natural phenomena, finding rules and paths where there seems to not be any. I agree that it is a tremendous hard work and, unfortunately, fallacies are the order of the day. Nevertheless, I firmly believe that, with the right tools, used with caution and deep understanding, science could be pushed forward and useful discoveries can be made.

Take as an example Autism Spectrum Disorder (ASD), as for other neurodevelopmental disorders, such as schizophrenia, its aetiology is not known yet. The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) defines ASD as a disorder characterized by (1) deficits in social communication and social interaction, and (2) restricted, repetitive patterns of behavior, interests, or activities, whereas atypical language development is regarded as a co-occurring condition. The diagnosis usually occurs in the first 3 years of life. The characteristics of ASD underwent several changes through time and now ASD indicates a complex multi-factorial disorder. On the one hand, genetic studies of ASD have identified mutations that interfere with typical neurodevelopment in uterus through childhood. On the other hand, neuroimaging studies have provided many important insights into the pathological changes that occur in the brain of subjects with ASD.

As so far discovered, ASD is characterized by phenotypic and genetic heterogeneity. Furthermore, ASD behaviors are not static within individuals across development. It is thus problematic to link

cross-sectional behavioral data to genetics, since findings of relationships become dependent on the specific point in time at which the behavior is measured. Using trajectories of change as behavioral phenotypes could be the key to provide insights into ASD aetiology. And this is where problems begin… Firstly, large longitudinal databases are needed, which should be able to provide evidences for the several phenotypic characterizations of the autism spectrum. Moreover, genetic information should be available as well. Secondly, analytic tools should be developed to actually analyze these data, providing pattern classifications and clear connections to genetic disruptions. To address the study of such a complex disorder, I believe that interdisciplinary research is needed.

Given my interest in the application of mathematics, I firstly got in touch with the Psychology and Cognitive Sciences Department (Rovereto, University of Trento) when I was searching for a topic for my master thesis. During that time I began to acknowledge the challenges related to Autism Spectrum Disorder and I committed myself to it since. Currently, I am a PhD student in Psychological Sciences and Education (University of Trento) with a grant from Fondazione Bruno Kessler (FBK). In this way I have the possibility to merge the two disciplines that need to "communicate" to shed some light on the ASD mistery: psychology and mathematics. On the one hand, I am learning how behavioral data are collected and interpreted from the collegues at the Laboratory of Observation, Diagnosis and Education (ODFLab) in Rovereto. Which diagnostic instruments are used to diagnose and assess ASD? What kind of interventions produce specific behavioral and developmental outcomes for individual children with ASD? On the other hand, at FBK (Predictive Models for Biomedicine and Environment), I am studying machine learning and deep learning techniques to develop a novel and original approach for the integration of phenotypic and genotypic data. The aspects related to computational biology will be supervised by the Center for the Integrative Biology (CIBIO, University of Trento).

I have already spent two years working with psychologists. We have different backgrounds, we talk differently, we address the same problem becoming aware of different things. However, step by step, differences can be overcome. Moreover, it became clear to me that the more I struggle with this new subject of study, the more I will be able to approach the study of ASD from a new angle, which includes the clinical knowledge of the disorder. Unfortunately, to receive an interdisciplinary education is not the only important step towards new scientific discoveries. As I mentioned before, large databases (in particular longitudinal databases) are necessary, and the collaboration with researchers with the same goal can increase the probability to find the genetic causes to ASD, reducing fallacies.

Given recent discoveries, I believe that a mathematical approach can help to discover the genetic factors underlying the different phenotypes characterizing ASD. In the future, I hope it will be possible to individualize treatment for ASD, finding the best therapeutic options. In order to do that, a rigorous interdisciplinary approach should be built, with a particular attention to its replication.

PERMALINK

https://magazine.fbk.eu/en/news/what-if-data-is-all-we-have/

AUTHORS

- Isotta Landi