Research Statement

Enrico Bertini
Associate Professor
NYU Tandon School of Engineering
May, 2021

My research interests lie at the intersection of data visualization, human-computer interaction and machine learning. The main goal of my work is to help people understand complex phenomena through the use of interactive data analysis and visualization tools. The guiding question of my research is: "*How can we develop interactive technologies that expand human capabilities in reasoning and communicating with and through data?*"

In recent years, we have witnessed an explosion of computational capabilities to help people reason and communicate with data. By analyzing large quantities of data from disparate areas, such as economics, finance, biology, medicine, we have the opportunity to better understand the world and potentially enable remarkable discoveries. Computational methods developed in databases, statistics and machine learning, enable us to process ever larger quantities of data and to build models in support of prediction, decision making and knowledge discovery. But, in order to harness this enormous power, it is crucial to develop theories, models and technologies that support human thinking in addition to automation. While a substantial proportion of methods aims at full automation, there is a strong need for technologies that support human thinking when and where it is needed. To this purpose, it is necessary to make progress in developing a productive dialogue between humans and machines. Innovative methods are needed to: (a) express intent and transfer knowledge from the human mind to the machine and (b) to enable effective interpretation and communication of the output generated by computational methods.

The research developed in my lab aims at bridging these gaps through the study of visual analytics interfaces: interfaces that enable reasoning with data through interactive graphical representations. This research is characterized by the development of innovative technologies as well as experimental work to understand how data interfaces support and expand human cognition. Most of our projects are carried out in collaboration with domain experts to ground our work on real-world problems and they often involve interdisciplinary teams including experts in machine learning, data visualization and psychology. In the following I describe in more details specific lines of research we have been pursuing in recent years.

Visual Interfaces for Interpretable Machine Learning. The recent advancements in machine learning (ML) and Artificial Intelligence (AI) have made the use of complex models ubiquitous in our society. ML models are used to make all sorts of crucial decisions that affect both organizations and individuals. For instance, ML models are routinely used to make decisions in healthcare, law, banking, finance, security. But, as models become more ubiquitous and consequential, it is also necessary to increase our ability to inspect, validate and understand them. In a recent interview study we conducted with data scientists from industry, we found that the problem of making models more interpretable affects many different stakeholders in an organization and that several research and technological gaps exist [4]. For instance, one clear need is the availability of model inspection and communication tools that permit data scientists to communicate and discuss model behavior with domain experts. Similarly, ML practitioners need tools that support capturing domain knowledge from experts as a way to debug and perfect existing models.

Regarding the development of interfaces that permit to inspect and communicate model behavior, we have recently developed a series of techniques and research prototypes to study the problem of model sensemaking. The main focus of these solutions is to provide support in gaining a *global* understanding of how a model behaves. For example, when inspecting a model, it should be possible to understand what data features and values drive specific decisions; whether the decisions make intuitive sense; and when (and possibly why) the model fails to provide reliable recommendations.

To attack this problem, we followed so far two main directions: *aggregation of instance-level explanations* and *global surrogate models*. Both solutions employ a "model agnostic" (post-hoc) approach, that is, rather than leveraging the internal logic or structure of the model, they try to infer the model's logic by inspecting its input-output behavior. The main advantage of such methods is that they can adapt to any model and as such they are very flexible. An obvious disadvantage is that they may not provide a faithful representation of the actual logic used by the model.

Aggregation of instance-level explanations leverages the many existing methods that have been developed to generate explanations of individual model decisions. Such methods typically generate sets of weights that describe relevance and role of a feature for individual instances. In our work, we use these explanations of *local* behavior to guide the user in inspecting the *global* behavior of the model. In *Rivelo* [9], we aggregate and visualize counterfactual explanations to understand the role individual and groups of words play in driving model predictions. In *Explanation Explorer* [5], we use a similar technique to understand predictive models developed at the NYU Langone hospital to predict hospital admissions. The system aggregates and visualizes large quantities of individual explanations to help data scientists and physicians at Langone validate the decisions the model makes. Our solution allowed the experts to review their models, identify important shortcomings and develop more effective solutions for delivering services in the hospital.

Global surrogate models use a different logic. Instead of starting from individual explanations, these methods aim at developing a description of an existing (black-box) model M_b by building a more transparent surrogate model M_s that simulates the original model M_b. The basic idea is to use the output of M_b as training data to build M_s, so that the more transparent model obtained by this process could be used as a description of M_b. Common approaches in this area involve rules and decision trees, which are commonly considered more intelligible and transparent than more complex models.

In this area we have so far developed two solutions: *RuleMatrix* and *SuRE*. *RuleMatrix* [6] generates surrogate rule sets and visualizes them through a novel visual representation based on a matrix metaphor. The matrix visualization has two main advantages: it makes it easier to visualize a larger number of rules and, more importantly, it aligns the rule predicates in a way that permits to quickly identify how features are used across the rules; a solution that enables better mental inference of how the model works. This work also starts exploring the trade-off between *fidelity* (how accurately the surrogate model mimics the original model) and *complexity* (how complex rules need to be to describe the original model faithfully). Initial experiments we developed show that the gain in fidelity obtained by producing more complex rules may not be justified given the limited increase in fidelity.

SuRE (Surrogate Rules Explorer) is a follow-up work that solves a problem we have identified while working on *RuleMatrix*. One major limitation of rules sets is that rules do not permit to understand what the role of individual predicates is for a given prediction. For instance, in a rule like the following: IF $f_i > a$ and $f_j < b$ THEN L (where f_i and f_j are data features and L is the model output), it is not possible to assess the impact of each of the individual predicates f_i and f_j on the prediction. A potential solution to this problem is to use decision trees, which permit to represent the role of individual predicates (nodes of the tree), however decisions trees have limited flexibility in representing complex decisions spaces and they can easily grow very complex. *SuRE* uses a lattice structure as a solution and a novel network-based visual representation which permits to observe the impact of individual predicates while maintaining flexibility and keeping complexity low.

Much more work is needed in this space. The solutions we developed so far work almost exclusively with tabular/structured data and do not immediately adapt to other relevant data types such as time-series, documents and images. Also, much more empirical work is needed to understand how these methods impact the experts' work when used in real-world settings. In this regard, part of our current effort is devoted to studying the use of *SuRE* in a team of data scientists at a large bank; a team of chemists who are studying predictions of chemical reactions; and a private company that focuses on the development of interpretability methods for large corporations.

Mixed-Initiative Interfaces for Data Analysis. When performing data exploration, interactive systems leave the burden of deciding what part of the data space to explore and how to explore it completely to the end-user. While this is desirable to maintain a sense of agency for the end-user, manual exploration is cognitively taxing and can lead to overlooking potentially relevant segments of the data space. In this area, our work focuses on developing mixed-initiative methods that support users in their data exploration activities.

We are currently working on a tool called *SliceLens* which guides users in the exploration of training data and model behavior in machine learning. *SliceLens* helps users segment the data into data subsets (data cubes) so that they can inspect model behavior and data distribution in such subsets. The analysis is guided by a series of "interestingness functions" which drive recommendations about what combinations of features/variables to explore during the analysis. For instance, such recommendations can guide the user in discovering data subsets with a high number of errors or inconsistent predictions. *SliceLens* represents an initial attempt at understanding how computational guidance impacts data exploration and how it could provide support to the end-user without being intrusive.

Another area of data analysis and machine learning where a mixed-initiative approach can help speed up and improve the quality of the process is data labeling. Lack of high-quality labeled data is often one of the main bottlenecks when developing new ML models. To overcome this problem, organizations often use crowdsourcing platforms to label large quantities of data. This solution is however not viable when the data to be labelled is sensitive and/or when labeling requires expert knowledge that can only be found "in-house". To address this problem we are researching *interactive data labeling* solutions to help small groups of domain experts label large quantities of unlabelled data with few errors and high trust.

An early attempt to deal with this problem is presented in *ELA*, the "*Exploratory Labeling Assistant*" [3], a mixed-initiative interface that integrates data visualization and recommendations to guide users in the ideation and assignment of data labels. Our experimental results show that users can efficiently label large collections of documents with high confidence and accuracy. Ongoing work in our lab is extending the work done with *ELA* to study how data labeling systems can leverage the user's knowledge more directly and how such interventions can help speed up the process while keeping high quality in the labeling task. We are also looking into how to make the labeling process more transparent by creating data visualizations that show the impact of labeling decisions on model training.

Visualization Literacy and Visual Communication. The recent widespread success of data visualization has led to the consumption of graphical information and statistics by a much larger segment of the population (e.g., through visualizations published in popular news outlets).

As more people consume information and make decisions based on data visualizations, it is crucial to better understand how such information effects people's perception and opinion of current events and facts. While extensive research in visualization has addressed problems of low-level perception, there is comparatively less research on how people extract messages and interpret information from graphical representations of data.

In this space, we have published a series of landmark papers that paved the way to the study of "visualization psychology". Some of this research has been spearheaded by a highly rewarding and successful interdisciplinary collaboration between our lab and a group of Human Rights experts from NYU Law. Rights activists can potentially gain many benefits from visualizations that depict data about human tragedies. Such representations may make their messages more transparent, impactful, and ultimately more persuasive. In a first paper we published together, we investigated the persuasive effect of visualization when compared to tabular presentation of data and found that visual depiction of data can lead to more pronounced persuasive effects [7]. In a follow-up work, we investigated the problem of *visual deception* and verified that commonly known graphical distortion techniques are largely overlooked by the readers and, as such, can easily mislead (a very alarming result) [8]. In a more recent follow-up work [2], we studied the effect of "axis truncation" in charts, a common technique used to artificially exaggerate the difference between data values. In a subsequent paper, we conducted experiments to verify whether visualization can help elicit more empathy in readers and, as a consequence, increase donations to a selected set of human rights causes [1]. Our studies repeatedly failed to find any effect of visualization on empathy and as such they challenge the commonly held belief that visualization can be of help in making data and statistics more "empathic" (a result that has been corroborated by several other researchers in subsequent studies).

More recently, we started a new collaborative project with colleagues at Merced and Georgia Tech to understand the impact of data visualization in communicating information about the covid19 pandemic. Understanding how such information affects people's understanding, emotions and behavior, is clearly very relevant. In an initial study we developed, we analyzed a large collection of covid19 visualizations to understand how information about the pandemic is currently communicated [10]. Ongoing studies we are developing are shedding a light on how people react to different types of visualizations and how different ways to communicate uncertainty and risk can lead people to different opinions and behaviors.

Much of the work we developed in this context is at the very forefront of our understating of how visualization impacts people's opinions and understanding of facts; it also bears practical implications for the use of visualization in settings that can potentially affect millions of individuals. Data visualization research and design have been guided for many years by a narrow focus on low-level visual perception. Our work demonstrates that there is much more to visualization than accurate extraction of quantitative values from graphical representations. We are currently working on developing a new theory for data visualization based on what we call "graph affordance", that is, graphical properties that better predict how people interpret information represented in graphs. With this research we aim at developing a new understanding of how visualization works and at providing better guidance for visualization design.

References

[1] J. Boy, A. V. Pandey, J. Emerson, M. Satterthwaite, O. Nov, and E. Bertini. Showing People Behind Data: Does Anthropomorphizing Visualizations Elicit More Empathy for Human

Rights Data? In *Proc. of the ACM Conference on Human Factors in Computing Systems (CHI)*, 2017. **Best Paper: Honorable Mention**.

[2] M. Correll, E. Bertini, and S. Franconeri. Truncating the Y-Axis: Threat or Menace? In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1–12, 2020. **Best Paper: Honorable Mention**.

[3] C. Felix, A. Dasgupta, and E. Bertini. The Exploratory Labeling Assistant: Mixed-initiative Label Curation with Large Document Collections. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, pages 153–164, 2018.

[4] S. R. Hong, J. Hullman, and E. Bertini. Human Factors in Model Interpretability: Industry Practices, Challenges, and Needs. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–26, 2020.

[5] J. Krause, A. Dasgupta, J. Swartz, Y. Aphinyanaphongs, and E. Bertini. A Workflow for Visual Diagnostics of Binary Classifiers Using Instance-level Explanations. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST)*, volume 4, 2017.

[6] Y. Ming, H. Qu, and E. Bertini. RuleMatrix: Visualizing and Understanding Classifiers with Rules. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VAST)*, 25(1):342–352, 2018.

[7] A. V. Pandey, O. Nov, A. Manivannan, M. Satterthwaite, and E. Bertini. The Persuasive Power of Data Visualization. *IEEE Transaction on Visualization and Computer Graphics (Proc. of InfoVis)*, 20(12):2211–2220, 2014.

[8] A. V. Pandey, K. Rall, M. Sattarthwaite, and E. Bertini. How Deceptive are Deceptive Visualizations?: An Empirical Analysis of Common Distortion Techniques. In *Proc. of ACM CHI Conference on Human Factors in Computing Systems (CHI)*, 2015.

[9] P. Tamagnini, J. Krause, A. Dasgupta, and E. Bertini. Interpreting Black-box Classifiers Using Instance-level Visual Explanations. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics (HILDA)*, pages 1–6, 2017.

[10] Y. Zhang, Y. Sun, L. Padilla, S. Barua, E. Bertini, and A. G. Parker. Mapping the landscape of covid-19 crisis visualizations. *Proc. of ACM Conference on Human Factors in Computing Systems (CHI)*, 2021.