Position Paper: Towards Comprehensible Predictive Modeling

Michael Gleicher, *Member, IEEE*

Abstract— Predictive modeling applications have a number of concerns. While predictive accuracy is often the primary concern, other issues such as efficiency and robustness are often considered. It is my position that comprehensibility is another concern in predictive modeling applications, and is often an important one. However, without an understanding of comprehensibility in predictive modeling, it is difficult to develop techniques to support it, or to make tradeoffs that balance it with other concerns. In this paper, I will provide thoughts as to understanding comprehensibility across the predictive modeling process. The initial framing of considering *who* is comprehending, *what* are they trying to comprehend, and *why* are they trying to comprehend it allows us to see a range of possible problems and answers to *how* we might better support predictive modeling applications. These initial thoughts suggest the importance of considering comprehensibility across the predictive modeling process.

1 Introduction

Predictive modeling applications consider a variety of factors. While predictive accuracy is often a primary concern, other issues, such as scalability, efficiency, robustness and verifyability, often must be considered. In any particular application, different factors may be important, and different tradeoffs between desirable properties can be made. By understanding the many desirable properties of predictive modeling applications, the various communities are making considerable progress at providing an arsenal of techniques that not only provide these properties, but provide different tradeoffs between properties so that appropriate balances can be provided to specific applications.

Comprehensibility, broadly the stakeholder's ability to understand, is another useful concern in the predictive modeling process. Like other properties, it may be more or less important in different applications. While there are some situations where a mysterious black box that makes accurate predictions may be acceptable, often understanding parts of the process of building and using that predictor can bring benefits. Also like other properties there may be tradeoffs, for example using a more complex (and difficult to understand) technique may lead to better accuracy at the expense of comprehensibility. However, comprehensibility seems to be a multi-faceted property of predictive modeling, relevant to the entire process in many ways. It is not well defined or easy to measure.

It is my position that we if we can develop a better understanding comprehensibility across the entire modeling *process* we will better be able to develop predictive modeling tools, techniques, and applications. It is important to understand comprehensibility so we can develop methods for achieving it, especially in light of the likely tradeoffs with other desireable properties. By taking a human-centric view of the process, we can identify challenges and opportunities for developing improved methods.

In this paper I offer some initial thoughts on how we might understand comprehensibility in predictive modeling. I suggest we consider three simple questions: *who* is doing the comprehending? *what* are they trying to comprehend? and *why* are they trying to comprehend it? I believe that these questions are can be considered independently, allowing us to mix-and-match answers appropriate to different sitations. Exploring these three questions is a starting point to answering the fourth question, *how* do we help them do it?

I emphasize that this is an *initial* attempt of formulating our understanding of comprehensibility in predictive modeling. I do not claim, or even believe, that it is complete, provides well defined (or even well-named) categories, or superior to other ways of organizing our quest to understand comprehensibility. I have found it a useful way of thinking about the problem space. For example, in identifying new opportunities for methods (hows) in unexpected combinations of whos, whats, and whys.

• *Michael Gleicher, Department of Computer Sciences, University of Wisconsin - Madison. E-mail: gleicher@cs.wisc.edu.*

Proceedings of the 2014 IEEE VIS Workshop on Visualization for Predictive Analytics. Available online at: `http://predictive-workshop.github.io`.

2 Who is trying to comprehend?

There are many potential stakeholders in a predictive modeling application. This range of stakeholders all have different needs for comprehending parts of the prediction process. As these needs are different, a sense the stakeholders is valuable in identifying how to provide them with appropriate support. While sometimes individuals may have multiple roles in the process, there are also times when different stakeholders have different abilities. For example, one might not be able to assume statistical sophistication in a general audience (for example, if a prediction is to be published in the popular press). Also, in practice there might not be clean separation between roles.

The stakeholders in a prediction application may include:

Developer: people building general technologies — for example a researcher inventing a new algorithms or implementations.

Data Scientist: people who do the model building. This role involves using the general tools, in a way that is somewhat detached from the domain.

Domain Expert - people who "have the problem," commission the model building, and consume of the results. In visualization, we often refer to such people as "domain collaborators."

Audience: people who ultimately get the results. For example, the audience of a scientists' papers or journalists' article.

3 What are they trying to comprehend?

Predictive modeling is a process. Arguably, the process begins before data gathering and wrangling, moves through the phases of mathematical model building and validation, and doesn't end until at least the users have had a chance to act on the predictions made. And of course, it may not end there, as reflection, reuse, and revision make the process potentially cyclic. Comprehensibility can be a factor in any of these phases in the process. This means that comprehensibility can mean many different things.

One aspect of the path towards comprehensible modeling is to identify and catalog the many kinds of comprehensibility that are relevant to the predictive modeling process. As an initial step towards this path, I offer a set of five different kinds of comprehensibility in the modeling process, organized by the phase of the process they are most closely associated with.

Inputs: Does the stakeholder understand the data, assumptions, and initial questions being used to build the model? For example, we might have a collection of training data or a causal theory that leads to a mathematical model.

Method: Does the stakeholder understand the method used to build the model from the data (e.g. the algorithms)? For example, in a linear support vector machine learning approach to convert the data into a predictive model, the method comprises both the general method (linear SVM) and the specific implementation of that method (e.g. a particular machine learning software package). In creating a model from some underlying principles, the method might be a symbolic derivation.

Model: Does the stakeholder understand the model used to make predictions? For example, this often takes the form of the specific equations that compute predictions. For the SVM example, the model is the set of coefficients (the linear equation) created.

Output: Does the stakeholder understand specific predictions made by the model? The output is the prediction computed for a specific prediction. It may include more than the prediction itself, for example, often predictions include confidence values, strength scores, uncertainty quantification, or sensitivity analyses. Output comprehension includes both the semantics of the result (e.g. what does the prediction of a specific value for the predicted variable mean) and the mathematical meaning of the prediction. For a common example, a weather prediction ("20% chance of rain tomorrow") may say little about any specific time in the day, the amount of rain that might fall, or whether it will be sunny.

Experiments: Does the stakeholder understand the results of experiments run on the model? Experiments include predictive performance evaluation (e.g. cross-fold validation or holdouts), and efficiency test (e.g. profiling). The results of experiments can include the outputs over a set of cases (often with known ground truth), but also various statistical information.

4 WHY ARE THEY TRYING TO DO THIS COMPREHENDING?

There are many reasons why understanding an aspect of the prediction process may be important. Or, put another way, there are many applications of comprehensibility. In generating an initial list, it is clear that there is a wide range, and that organizing this range will be a challenging but useful task. In generating this list, I have sought to define reasons that are independent of the specific things being comprehended, and generalize a range of possible pathways towards achieving the goal.

Build Pre-scriptiveness: understanding how to make use of the prediction. For example, to use an extended weather forecast to choose appropriate clothing.

Build Trust: understanding a prediction to appropriately trust it, or a predictive process to trust in its ability to make predictions (and to trust those predictions).

Improve Performance: understanding can lead to improvements in many of the other desired properties. Iterative refinement is often applied to improve predictive accuracy, efficiency, robustness, and even comprehensibility.

Discover Causality / Build Theory: While the predictive modeling process typically has its main goals of making predictions, it can often have the side effect of shedding light on the underlying process that is being modeled. While this usually means finding correlations, not really identifying causality, it can be a useful starting point for theory building, or even an empirical approach towards testing theory.

Extend / Characterize: understanding what the model can do and where it can be applied. For example, even if models do not have an explicit uncertainty characterization for their predictions or an explicit characterization of their operating regions, a deeper understanding of various aspects of the model can help create them.

Generalize: understanding can help extend modeling work to situations beyond its original goals, for example to see that methods might apply in other applications.

5 HOW TO HELP?

The important question is how to best support the stakeholder who is trying to understand the aspect of the process for the reason that they have. An observation from the early exploration is that a variety of combinations of these three are possible. Indeed, my own experience has been that trying to support an expected combination often leads to unexpected ones. Here I provide a two examples from my recent work. These examples are meant to give a sense of the ways these unexpected combinations occur - there are certainly mainly other examples out there in the works of others.

Molecular Surface Binding Experiment Visualization: In [2] we describe a system designed to help in understanding the results of validation experiments for machine learning methods designed to predict protein binding locations. We initially expected the system to be useful for the developers / data scientists who were trying to understand the performance of their prediction systems in order to improve them. As we started to deploy the system, we found it served a number of other ends. For example, we found that the system was interesting to the domain scientists as it helped them build trust in the model. It also turned out to be useful in building theory of the underlying mechanisms: seeing groups of proteins that were hard for the prediction suggested that there were alternative (biochemical) mechanisms involved in some cases. The domain scientists even used the visualization of the experimental results to communicate with their audiences.

Explainers: In [1] we describe an approach to exploring high-dimensional data by constructing classifiers. The goal (why) of these classifiers is to organize the data or suggest relationships between variables to help in building theory of underlying causality. Therefore, the metrics for choosing "good" classifiers are how well they serve these goals, not necessarily the predictive accuracy of the classifiers. The approach introduces methods for making tradeoffs between accuracy and properties thought to better enable theory building and organization, such as sparsity, simplicity, and diversity.

6 LESSONS

Comprehensibility is often an important concern for predictive modeling. In order to begin to develop tools that help support understanding prediction, we must improve our understanding of this understanding. There are opportunities for visual analytics tools and human-centric thinking to improve comprehensibility for many different types of stakeholders, across the many stages of the modeling process, and for many different goals.

ACKNOWLEDGMENTS

This paper has grown out of discussions with many collaborators, including my students, my domain collaborators, and everyone else who will listen to me. This research was supported in part by NSF Awards 1162037 and 0941013, and by a grant from the Andrew Mellon Foundation.

REFERENCES

[1] M. Gleicher. Explainers: expert explorations with crafted projections. *IEEE transactions on visualization and computer graphics*, 19(12):2042–51, Dec. 2013.

[2] A. Sarikaya, D. Albers, J. Mitchell, and M. Gleicher. Visualizing Validation of Protein Surface Classifiers. *Computer Graphics Forum*, 33(3):171–180, June 2014.