Question 9

The Rashomon set is an available problem for the area of machine learning. Although it may sound reasonable to start using the Rashomon curve and ratio as a method to select the most explicit method, it is not such a useful reliable tool as the authors considered.

The work done by Dr. Semenova and Dr. Rudin make an attempt to persuade the reader to avoid using the Blackbox in high-stake decision making or trouble shooting models and use the inherently interpretable model. This judgment established on some deficient aspect competing to transparent model, including high complexity, huge calculation, without faithful interpretable model, incompatible to additional information, etc. As some of the cases that stated in the paper, predictions made in industry like financial, health care, criminal justice prefer an explainable model than an accurate result. However, those points are primary based on the unexplained attribute of the popular machine learning models, and left out two main points: all models are uncertain and time-sensitive, and the necessarily of applying an interpretable model.

Firstly, models are risky and sensitive to time and situation. That applies to both unexplainable models and explainable. One of the problems of the most machine learning model is the variability of model selection and the high-dimensional dataset. These typically can be reduced by techniques like ensemble and boosting, and constant testing and updating. The transparent model also diverse in a way that different analytics have their own way of modeling constructing and setting criteria. Especially for high-stake decision problem, because the more human interference, the higher the require for knowledge, the more likely it go wrong. Our human is much more unreliable than the computer, and that could be the benefit of unsupervised learning. For instance, the machine learning project built by Youyang Gu for recent epidemic prediction model exceeds the ones made by the professionals. Without additional manually interference, the number of deaths could provide sufficient information than other countless factor that professionals may value.

I admit that relative to machine learning, explainable models are controllable and flexible, and a simpler model could achieve similar performance. There is not necessarily to pursue a totally transparent model. Interpretable is a vague descriptive tool that varies according to the audience and the interpreter. Depending on the needs of the audience, it can be used as one of the cutting points to evaluate the model. But for models that does not relates to subjective topic like the criminal behavior, there is no needs for a complete transparent model. Sometimes, the number itself persuade better to the receiver than a persuasive talk or technical judgement.

To summary above, the desire for data interpretability is based on 1. adequate resources (large data with strong correlation/causation, time, human and technical expertise); 2. minimal human interference-impartiality; 3. (importantly) the problem itself has a need for model reliability (e.g., the need

for fault detection).

As for concept of Rashomon effect and Rashomon set, it is more a presumption than a proof. According to Fisher et al. (2019), Rashomon sets summarize the range of effective prediction strategies that an analyst might choose. Additionally, even if the candidate models do not contain the true data generating process, we may hope that some of these models function in similar ways to the data generating process. By certain valuation, some explainable model could be capture. Nevertheless, it is hard to put models with different decision-making mechanisms in the same table to make a comparison. Besides that, although it is reasonable to assume the existence of relative more explicable model, the presence of non-empty Rashomon set is not yet proved. Therefore, other than logic, the idea of Rashomon set is not mathematical verified, that reduced the credibility of this method. In addition, the Rashomon set broaden the range of candidate models, which may introduce more subjective opinion and provide some freedom to the practitioner who might be exploited in a harmful way.

The concept of Rashomon introduce the concern of the correctness of models, and the necessarily of model transparency. Since all models are wrong and inaccurate, and our general purpose is to find the most predictable and useful ones, the Black boxes are applicable enough. And just like the name of Rashomon effect, ‘Nothing in our definition of interpretability hints that a model will ultimately meet its requirements in terms of interpretability.’.

References

Lou Y, Caruana R, Gehrke J. Intelligible Models for Classification and Regression. In: Proceedings of Knowledge Discovery in Databases (KDD). ACM; 2012.

Semenova L, Parr R, Rudin C. A study in Rashomon curves and volumes: A new perspective on generalization and model simplicity in machine learning; 2018. In progress.

Cynthia Rudin, Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. stat.ML 1811.10154; 2019.