

The Amalgamation of Deep Learning Model and Conventional Statistical Model

Hao-Wen CHI

*Kang Chiao International School, Kunshan, China
qihaowen20262023@163.com*

Abstract: With the development of current society and the expanding population, deep learning models start to show their advantages over other statistical models--especially the conventional statistical model (CMSs), like the applications of normal distribution. This research mainly discusses the pros and cons of deep learning model compared to conventional statistical models and how to take advantage of both models to obtain a more comprehensive model. In the current society, with the rapid scientific development, research often requires serious, precise, and accurate data and analysis to achieve perfect results. This is what researchers want to see. Ideally, when experimenters amalgamate these two models--CSMs and deep learning model- some data analysis will become more precise than people previously thought. This is the purpose of this research wanted. By observing the pros and cons of these two models and making a comparison between them, finding out the previous cases of applications of the specific model used, as well as the analysis of the data provided by the references, this research finds that, when people take both advantages of these two models, the results offered by these models will be more accurate and less biased, regardless of whether the prediction of the data or the regression line is used to find out the relationship between the x value and y value.

Keywords: deep learning model, conventional statistical model, prediction, bias

1. Introduction

Deep learning is a model utilizing algorithms that use huge data references to establish classification, make predictions, and evoke inferences [1]. On the other hand, conventional statistical models are an approximation of the real system, restricted by the knowledge and understanding of the dynamics from real experimental observations, which have the major challenges for predicting a large number of parameters [2]. In the current society, with the advanced prediction made by computer programs, people can easily, and concisely get precise data from the database. For example, the experiment done on a specific type of virus, which causes a severe disease, was solved by machine-learning techniques that can produce models that are able to utilize a large amount of multimodal data to produce multiple predictions rapidly. This has been demonstrated by its promising use in this particular disease to produce ever more accurate predictions [3]. From the case studies, people can perceive how the machine-learning techniques are manipulated to make the research more convenient and rapid so that people can efficaciously prevent some catastrophic events from happening. From this research, the main purpose is going to find out how to utilize the deep learning model to make sure it is running with the best effort,

combined with conventional statistical models(CSMs), to prevent the coming of such a similar event, like another catastrophic virus, which infected people will suffer from. Accompanying with this strong and effective help from the models, some patients or other assembly of people can more efficaciously tackle the issues that people confront in the daily routine. This research mainly shows the difference between the manipulation of CSMs and deep learning models and some advantages and disadvantages of both the CSMs and deep learning model to find if there are some unexpected outcomes with this combination, to achieve more advances in all the fields which need the help of prediction.

2. The difference between conventional statistical models (CSMs) and the deep learning models

Nowadays, more and more people start to work with deep learning models, which serve as an innovation during the scientific research, while still some people learn and utilize the CSMs. With the help of distinct systems, the result provided usually will happen with a tiny discrepancies. Therefore, there is a problem emerging in our sight. According to many confirmations of references, there are 4 distinct discrepancies between them. The first discrepancy is the different method of manipulation. Talking about the CSMs, the method to approach is through the assumption of the linear regression model, to use the explanatory variable to gather the predicted variable from the line $y=ax+b$. The entire, complete procedure is, firstly, one must perform the ideal experiment several times in the scenario where there is no change in the factors contained. In addition, researchers gather these data and plot the points on the graph to ensure there is no plot error. Moreover, according to the graph, researchers can analyze the data and add a normal regression line used to make the relationship between the x value and the y value-normally a straight line. Finally, it is the step to analyze the regression line to find out the direction, form, strength, and outliers. On the other hand, according to LeCun et al., deep learning models are being defined in the way that a multilayer stack of basic modules, many of which calculate non-linear input-output mappings and all (or most) of which are learning, which are simply not identical to that of the methods of CSMs [4]. Subsequently, the difference between them is the different interpretations used in the models. For the CSMs, the results that people get from the experiment may not be as precise as people's thoughts. According to Professor Andrew J. Vickers, a professional biologist and methodologist, conventional statistical techniques for assessing prediction models don't reveal anything about their clinical utility. In general, metrics like p values, relative risks, and the concordance index are unable to determine which of two models is better, whether it is worthwhile to collect data on an additional predictor, or whether a model is even worth deploying. Although calibration lacks a clear summary statistic, it is a necessary but insufficient prerequisite for using a model [5]. But then, like many simple models, deep learning models can offer how-possibly explanations of a variety of occurrences, which serve as the initial stages in identifying which causal mechanisms or dependence linkages require additional investigation. Due to its distinction of accuracy, it is probable that people can work further on this program. The next discrepancy is the application of these two models in real life. Due to the fact of the causality of CSMs, people can put this model in some professional work fields, like some fields of infectious diseases. In the research on the disease mentioned above, CSMs can be used to make conclusions about the relationship between the two variables through their causality. On the other hand, although deep learning models are able to generate a model to fit the real scenario and to make the assumptions and predictions rapidly, this type of complex model does not meet the needs for researchers and disobeys the classical statistical analysis because the things that a computer program can only do are to generate the accurate outcome from the given data, not the relationship, cause-and-effect between two variables. As a result of that, CSMs are suitable for researchers to make predictions when there is a need to search

for relationships and causality. Otherwise, the rest of the program is suitable for the deep learning program. Finally, the eventual part is the difference in generalizations between the two models. For the CSMs, the range for its to make generalizations might be associated with the integral representation of the Jack polynomial-the one that can induce that vertex operators used in the CSM are the primary fields with dimension one- is concluded by professor H. Awata, the one who serves as a member of the Mathematical Society of Japan and has research interests in the representation theory of infinite dimensional algebras [6]. On the other hand, learning with deep learning models displays good generalization behavior in practice, a phenomenon that remains largely unexplained. The main questions that remain unanswered are how the optimization algorithm selection biases such complexity to be low and how exactly optimization and implicit regularization relate to one another. So the generalization of deep learning models is still an unsolved problem in the field of statistics.

3. The advantages and disadvantages of deep learning models compared to those of the CSMs

After talking about the difference between the CSMs and deep learning models, the necessary things that this research must do is to compare these two models, get the advantages from both models, override their drawbacks, and amalgamate the useful parts. Talking about the CSMs, the absolute advantage for it is the definite precision, though this model cannot conclude the relationship. If people have a model that is derived from $y=ax+b$, it is easy for the deep learning models to find out the exact values of a and b and use the expression to predict the possibility of one specific value. That is what CSMs cannot tackle. For example, in the experiment about the predictions of 6-month outcomes in pediatric patients sustaining traumatic brain injury performed by Professor Andrew T. Hale, using an ANN, they developed a robust model in this work that can accurately forecast the outcomes of kids with traumatic brain injury (TBI). They believe that using this model can significantly enhance result forecasts for the assessment of these injuries in the scenario where they utilize a different model to predict the values. Evidence-based prediction software, like the data presented here, is set to become commonplace in the modern healthcare environment as computer and smartphone access grows worldwide and electronic medical records advance. The researchers show the superior accuracy of the tool, which is the first clinical prediction tool to use deep learning models to predict 6-month post-TBI outcomes in children. That is what the CSMs cannot deal with. In the second case, according to the experiment performed by Sheojung Shin, the professor compares the data expressed by the deep learning models to that expressed by CSMs about the research topic-the prediction of heart failure (HF) [7]. Due to the fact that the results made by deep learning models tend to have more accuracy than CSMs, the algorithms from deep learning models have the trend to become more different, which is even more perfect, than the algorithms from CSMs. As a result, the professor strongly suggests making his study with the method of deep learning models after he calculates the calibration and the readmission of his study because deep learning models have the approach to predict the risks of the readmission of the simulation and the mortality in heart failure (HF) patients. This can be defined as the second interpretation of accuracy. Nevertheless, there still are some needs for the deep learning models to meet, like the extrinsic justification. That is the reason why people need to amalgamate deep learning models with the CSMs since the CSMs can offer some correct justifications and validations for the needs of deep learning models.

4. Conclusion

This research discusses the different manipulations of two models with some discrepancies. In addition, this research asserts the many of the pros and some of the cons of the deep learning models. From these parts of extent, this research finds out that the utilization of CSMs is only available in the field with the demands of relationships and associations, though it still can make predictions, that are not very accurate. On the other hand, the deep learning models can make precise results, even with no errors. As a result, CSMs should be appropriate in the field of medicine, which can provide specific relationships and associations with one specific patient who suffers from an ailment, or even some cancers. Talking about this research, there are still many things that can be ameliorated, for instance, the results that are being predicted by the combination of CSMs and deep learning models in real life. However, there is, currently, no apparent case that uses this combination in real life. So, this part is what humans can focus on in future research and experiments, to find out this combination, even a model that can replace this combination.

References

- [1] Sullivan, Emily. "Understanding from machine learning models." *The British Journal for the Philosophy of Science* (2022).
- [2] Arun Pankajakshan, et al. "Experimentally Driven Guaranteed Parameter Estimation: a Way to Speed up Model-Based Design of Experiments Techniques" (2018) <https://www.sciencedirect.com/science/article/abs/pii/B9780444642356500656>
- [3] Al-Hindawi, Ahmed, et al. "COVID-19 prognostic models: a pro-con debate for machine learning vs. traditional statistics." *Frontiers in digital health* 3 (2021): 637944.
- [4] Sullivan, Emily. "Understanding from machine learning models." *The British Journal for the Philosophy of Science* (2022).
- [5] Vickers, Andrew J., and Angel M. Cronin. "Traditional statistical methods for evaluating prediction models are uninformative as to clinical value: towards a decision analytic framework." *Seminars in oncology*. Vol. 37. No. 1. WB Saunders, 2010.
- [6] Awata, H., et al. "Collective field theory, Calogero-Sutherland model and generalized matrix models." *Physics Letters B* 347.1-2 (1995): 49-55.
- [7] Shin, Sheojung, et al. "Machine learning vs. conventional statistical models for predicting heart failure readmission and mortality." *ESC heart failure* 8.1 (2021): 106-115.