RELIABILITY OF MACHINE LEARNING MODELS

Seval URERSOY

May 2022

1. INTRODUCTION

Machine learning is the ability of the computer to learn from train data or examples and perform to model building tasks for associated problems. Machine learning has become a significant competitive differentiator for today's world since machine learning models have been increasingly used in many fields. Especially big companies are making machine learning a core part of their operations and decision strategies. Not only in business but also in healthcare, social media, e-marketing and so on are the fields that Machine Learning is commonly used. Being ensure about the reliability of machine learning algorithms which are getting widespread in every field is critical issue for users. Reliability of machine learning models, under which circumstances, we can trust machine learning and what we need to be careful while applying these algorithms will be discussed in this paper.

2. HOW DO MACHINE LEARNING MODELS WORK?

Arthur Samuel described machine learning as a "Field of study that gives computers the ability to learn without being explicitly programmed" [1]. This definition shows how a machine learning model works in basic terms. Machine learning is set of algorithms which can be used for prediction, making inferences, understanding patterns or forecasting. Machine learning algorithms can be classified into four main categories such as unsupervised learning, supervised learning, semi-supervised learning, and reinforcement learning [2]. According to data and problem that we have, these categories and included different type of algorithms can be used. These algorithms basically learn by available data and perform given tasks. Each algorithm has it owns pros and cons and being aware of these is important steps for using Machine Learning models.

3. RELIABILITY OF MODELS

Before applying any machine learning models to available data, first step is to being aware of data quality. Possible problems may occur in model can be caused by insufficient data, unqualified data or not checked assumptions before modelling steps. Increasing data quality also increases model performance as well as reliability of future models[3]. Understanding data structure, checking data accuracy and finding suitable algorithms are good steps to conduct any models however one should be aware of the limitations in data sets every time. For example, unbalanced data sets in labeled data, missing values or high/low dimensions can affect model performance as well as reliability of models. Although one sure about data quality, it is not the only case in model reliability. Finding suitable algorithms and understanding drawbacks of these applications are really important to trust the model. For instances, Long short-term memory (LSTM) performed to estimate the health index of a system in Malhotra's study[4]. It has shown that although LSTM performs excellent efficiency and accuracy in prediction, it has overfitting problem. Overfitting can be misleading results and model performance since accuracy is very high in overfitted models. It might be confusing for success and cause work with unreliable models. Not only overfitting but also underfitting is one of the reasons why we can not trust immediately to machine learning models. When evaluating model performance, we use different types of performance metrics according to the model type such as F-Measure, Area Under the ROC Curve (AUC) and Root Mean Square Error (RMSE). Thus, using suitable metrics for the model and interpret model performance is the key to trust machine learning models. To be able to select performance metrics for conducted models, multiple performance metrics is suggested since using only one metric can be insufficient. Choosing multiple metrics that are not highly correlated with each other might be helpful for evaluating model performance as well as understanding how much we can trust machine learning model [5].

4. CONCLUSION

Machine learning models are critical for decision making and predictions for many fields that is why relying on machine learning models such an important criterion in data science. Different types of problems such as data quality, overfitting/underfitting or performance evaluation of models might affect reliability of machine learning models. We can not say that if these problems solved, models will become more trustworthy. However, being aware of these problems and

sevalurersoy@gmail.com

their possible solutions, we can create more reliable models. To conclude, we can not truly trust machine learning algorithms but understanding drawbacks and limitations in machine learning can create accurate and reliable models.

5. REFERENCES

[1] A. L. Samuel, "Some studies in machine learning using the game of checkers. ii—recent progress," 1967, vol. 11, pp. 601–617.

[2] Mohssen Mohammed, Muhammad Badruddin Khan, and Eihab Bashier Mohammed Bashier, *Machine learning: algorithms and applications*, Crc Press, 2016.

[3] Valerie Sessions and Marco Valtorta, "The effects of data quality on machine learning algorithms," in *ICIQ*, 2006.

[4] Pankaj Malhotra, Vishnu Tv, Anusha Ramakrishnan, Gaurangi Anand, Lovekesh Vig, Puneet Agarwal, and Gautam Shroff, "Multi-sensor prognostics using an unsupervised health index based on lstm encoder-decoder," *arXiv preprint arXiv:1608.06154*, 2016.

[5] Yangguang Liu, Yangming Zhou, Shiting Wen, and Chaogang Tang, "A strategy on selecting performance metrics for classifier evaluation," *International Journal of Mobile Computing and Multimedia Communications (IJMCMC)*, vol. 6, no. 4, pp. 20–35, 2014.