

Method of Performing Machine Learning

Nripesh Trivedi

Department of Mathematical Sciences,
Indian Institute of Technology, Varanasi

ABSTRACT

Currently, there is no one common measure to judge machine learning methods and results. Through, this paper, I introduce such a measure to judge quality of machine learning methods and results. We just measure machine learning algorithms based on accuracy of prediction. If there is a common measure to define quality of machine learning methods and results then it could be used in different situations like interdisciplinary research. Moreover, the results are just limited to the data that we have in hand and any change in data changes the machine learning methods and results. Therefore, to become independent of data a measure is required to show the quality of methods in a way that is independent of data. This may be the first paper to present such a measure. When a measure independent of data is present, it will help us in determining the applicability of machine learning methods and results in a way that is independent of data. Correctness is such a measure. When I say correctness, I mean justified mathematically.

Keywords

Machine Learning

1. INTRODUCTION

The method of doing machine learning should be that it should have results that are **correct** and that should be the only criteria. Instead of predicting the accuracy of outcome based on data, correctness of the result should be verified **mathematically**. We depend on data to verify the accuracy of the outcome or prediction. The method should be such that it should give the correctness of the result in that scenario. In this way, machine learning would advance to different fields based on the **correctness** of the result rather than depending on the just accuracy of prediction. By correctness of the result, I mean that it is **justified**. For example, an example of correct result is a result which satisfies an equation. An equation could have the form $-x^2 + 2x + 1 = 0$. This would mean that the result satisfies this criterion and is correct. Correctness is justified mathematically by the equation. Correctness would mean that these results could be used anywhere depending on the quality of the result which should be verified mathematically based on equations. The method of performing machine learning should be such that the result obtained should be correct and this correct result could then be used again for some purpose. The problem with the current machine learning methods and results is that they couldn't be used again due to lack of a measure that is common anywhere. Through this paper, I am proposing a measure that could be used to verify the applicability of the results of machine learning. This measure is **correctness**. By correctness, I mean **justified mathematically**. This is how the machine learning methods should be made as well. This measure should not just be limited to the machine learning results but should also be incorporated in machine learning methods. Therefore, one measure to judge the applicability of machine learning methods and results, that is correctness. As given before, an example of correctness comes from algebra in mathematics. An equation of the form $x^2 + 2x + 1 = 0$. could

be used to show the correctness of a result since it satisfies this equation and based on the correctness of the result, this result could be used at other places. Again, by correctness, I mean justification. Mathematics is the easiest way to justify something through equations and logic.

When a result is correct, it has the potential to be used anywhere again if it is applicable. Instead of verifying the accuracy of the outcome, the result should be correct mathematically so it could be used again based on the correctness. The correctness of the result means justification mathematically. It could be understood from the mathematical example given above and machine learning method examples given below. Any change in data changes the result, thus there is a need for a common measure to make the result applicable in various situations and events, therefore need of correctness.

Need of Correctness as a measure

Correctness as a measure is needed since there is no one common measure present to judge the **quality** and applicability of results and methods in machine learning. When a result is correct, it has the potential to be used somewhere else since it is correct and thus could be checked for its **quality** based on correctness. Moreover, a result when correct is useful in the sense that it could be checked for its applicability and quality. This measure makes our methods and results independent of data and would help in making the methods and results applicable to different type of events and situations. In [1], I show my method and results by using data. This is the common technique to show results as shown in [1] and [2]. If the data changes, the applicability of result changes therefore correctness would help us in making our result generalizable or simply in using somewhere else.

Example of correctness in present results

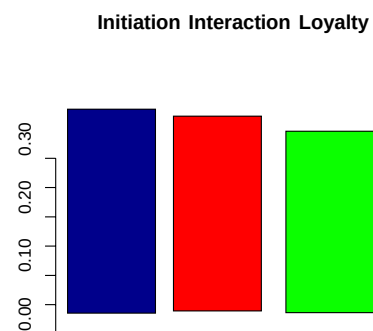


Figure 1: Category 2

This graph was obtained for the following class of people on a social networking website [2].

Category 2 Discussing and promoting malware, spyware

This means people showed a decreasing trend in terms of their interest in discussing and promoting malware and spyware [2]. This result is mathematically correct in terms of it being **true by sense**. That is, people actually are not interested in indulging in any activity that is of above category as mentioned in category 2 just by **sense**. Everyone loses interest in discussing and promoting anything bad such as spyware and malware by **sense** point of view. This result could be used somewhere else as it is generally **true** by just **sense**. Therefore, it is correct.

Example of correctness in present machine learning methods

Correctness in unsupervised learning

Correctness in unsupervised learning is about approximating the points with equations and then verifying the grouping of points by placing each point in the equations and checking it mathematically. Possible relations between points grouped together in different clusters could be found out as well by this way.

Correctness in Supervised learning

Correctness in supervised learning is about approximating the points with an equation in the training set and verifying whether the points belong to this equation among the points to be classified, based on training set. Possible relationships could be easily found out among points in this way.

2. CONCLUSION

Correctness (or justification) is a measure that is independent of data. Thus, it is very straightforward to use this measure to generate and use results that would further help us in performing machine learning.

3. REFERENCES

- [1] Trivedi, N., Asamoah, D. A., & Doran, D. (2018). Keep the conversations going: engagement-based customer segmentation on online social service platforms. *Information Systems Frontiers*, 20, 239-257.
- [2] Trivedi, N. (2017). Topic-Based Engagement Analysis: A Case Study.