Paper review [1]

Yu SU 55323707

1. The Problem

Machine learning classifiers are popular these years, while most researchers only use it as a black box without optimizing it deployment and configurations. Therefore, the paper tried to understand the performance of MLaaS systems and find the relation among the complexity, performance and the performance variability. What's more, the paper also tried to find the key knob of the performance and fixed it.

2. Challenge

The largest challenge is it is hard to give a fair evaluation on the performance and the complexity because only with a good evaluation criteria, the analysis is meaningful and correct. What's more, there are limited research on MLaaS platform, so it is hard to understand how the system works, so this is another challenge for doing an analysis on MLaaS system. And the other challenge is the past work on empirical analysis is only focus on a single step of the ML pipeline, but the paper tried to analyze end-to-end impact of complexity on classifier performance.

3. Key Insight

There are three main key insight in the paper.

First, the paper presented the correlation between the complexity and the performance. And found the highly tuned platform could work better such as the Microsoft, which means the more control dimensions for user the better performance it could get.

Second, although high tuned platform gives the higher upper limit, the paper found increased control lead to higher risk, which means if the user could not control the dimensions well, they would get a worse performance even without control. What's more, the paper found the users do not need to deploy the whole classifier and what they need to do is only focusing on three classifiers.

Third, the paper presented that the hidden optimization of the MLaaS system for the automated portions of the pipeline is great, though we could do it better by ourselves by quantitatively analyzing the optimization strategy.

4. Limitation

There are three main limitations for their studies. First, the research only focused on the target that supports large scale measurements, which means the result is maybe biased. What's more, besides the platform, the task and the configuration that they research on are also not generalizable. Last, the paper only focused on the relationship between the performance and the complexity. There are still more meaningful properties they should also care about, such as the scalability, cost and the robustness.

5. Future Work

As the limitation part said, the analysis of ML is not complete, so the future work is extending the analysis in the four dimensions, platform, task, configuration and properties to get a generalizable analysis result.

[1] Y. Yuanshun, X. Zhujun, W. Bolun, V. Bimal, Z. Haitao, Y. Ben(2017).Complexity vs. performance: empirical analysis of machine learning as a service. *Proceedings of the 2017 Internet Measurement Conference pages 384-397.*