Shangru Yi
GT - syi73
2020/01/14

Paper Review Week 1
TFX: A TensorFlow-Based Production-Scale Machine Learning Platform

CS 8803

1 Paper Overview

"TFX: A TensorFlow-Based Production-Scale Machine Learning Platform", written by Denis Baylor, Eric Breck, et al. from Google, is published in KDD 17'. Within this paper, they develop a TensorFlow-based all-in-one machine learning system to address the potential problems caused by diversified individual behavior and intrinsic of ML tasks. By identifying and coordinating the ML system components, they alleviate the possible technical shortages and burdens in system maintenance and updates. Combining with experienced problems in ML systems, this paper provides solutions to routine ML system problems, including data update, configuration and model validation [1]. Through a case study by applying their system to Google Play, an app store with large-scale ML recommending algorithm, they demonstrate the improved efficiency and effects of this integrated ML system, comparing with previous one in terms of several system aspects. To conclude, their contributions are summarized as follows:

1. The system flaws in traditional developing and maintenance mode of ML systems are identified and addressed by their published integrated ML system.
2. A case study regarding real-world situation with classical system level risks are studied and it demonstrates the improved effects by applying their developed system.

1.1 Problem Summary

Creating and maintaining a platform for reliably producing and deploying machine learning models can be particularly challenging with the feedback data changes over time [1]. Present systems are lack of unified configuration control, generating potential system level risks in update and maintenance. With relative large data scale, the models need to to be refreshed and produced continuously, which leading emerging risks for model training and deployment.

1.2 Related Works

- Hidden technical debt in machine learning systems [2].
- Keystoneml: Optimizing pipelines for large-scale advanced analytics [3].

2 Paper Strengths

The great efforts for providing an integrated and uniform machine learning system with the provided solutions are the main strengths of this paper. By combining with experienced case analysis and study, the possible improvement achieved by integrated system is demonstrated. The combined concrete thoughts based on TensorFlow is another highlights of the proposed system solutions. The case study of Google Play is great, given the technique background and data scale of that recommending system.

1. It is insightful to see the engineering concerns from Google.
2. The great efforts to address system flaws based on TensorFlow platform is concrete.

3 Paper Weaknesses

Although this paper provides solutions for the bring-up problems in ML systems, the solutions seem to more like an engineering experience, instead of technique quantified analysis. This paper is more like an experience sharing rather than an academic paper. Given the huge difference in technique background and system scale of Google, some small machine learning systems might not benefit from the proposed system frameworks or solutions. Also, since there is no proper metrics or detailed system behavior analysis, the strength of case study is lack of enough confidence.

1. The possible solutions for the diversified examples are experienced.

2. There is no available metrics given in this paper.

3. There is few detailed analysis for the provided case study.

References

[1] Denis Baylor, Eric Breck, Heng-Tze Cheng, Noah Fiedel, Chuan Yu Foo, Zakaria Haque, Salem Haykal, Mustafa Ispir, Vihan Jain, Levent Koc, et al. Tfx: A tensorflow-based production-scale machine learning platform. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1387–1395. ACM, 2017.

[2] David Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-Francois Crespo, and Dan Dennison. Hidden technical debt in machine learning systems. In *Advances in neural information processing systems*, pages 2503–2511, 2015.

[3] Evan R Sparks, Shivaram Venkataraman, Tomer Kaftan, Michael J Franklin, and Benjamin Recht. Keystoneml: Optimizing pipelines for large-scale advanced analytics. In *2017 IEEE 33rd international conference on data engineering (ICDE)*, pages 535–546. IEEE, 2017.