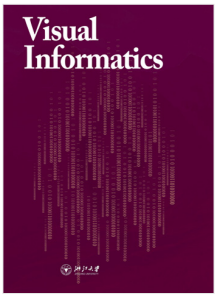

Towards better analysis of machine learning models: A visual analytics perspective

Shixia Liu, Xiting Wang, Mengchen Liu, Jun Zhu

PII: S2468-502X(17)30008-6
DOI: http://dx.doi.org/10.1016/j.visinf.2017.01.006
Reference: VISINF 6

To appear in: *Visual Informatics*

Please cite this article as: Liu, S., et al., Towards better analysis of machine learning models: A visual analytics perspective. *Visual Informatics* (2017), http://dx.doi.org/10.1016/j.visinf.2017.01.006.

Journal of Visual Informatics
ISSN-2468-502X (online)
www.evise.com/evise/jrnl/VISINF
E-mail: visinf@cad.zju.edu.cn

Towards Better Analysis of Machine Learning Models: A Visual Analytics Perspective

Shixia Liu*, Xiting Wang, Mengchen Liu, Jun Zhu

Tsinghua University, Beijing, China

Abstract

Interactive model analysis, the process of understanding, diagnosing, and refining a machine learning model with the help of interactive visualization, is very important for users to efficiently solve real-world artificial intelligence and data mining problems. Dramatic advances in big data analytics has led to a wide variety of interactive model analysis tasks. In this paper, we present a comprehensive analysis and interpretation of this rapidly developing area. Specifically, we classify the relevant work into three categories: understanding, diagnosis, and refinement. Each category is exemplified by recent influential work. Possible future research opportunities are also explored and discussed.

Keywords: interactive model analysis, interactive visualization, machine learning, understanding, diagnosis, refinement

1. Introduction

Machine learning has been successfully applied to a wide variety of fields ranging from information retrieval, data mining, and speech recognition, to computer graphics, visualization, and human-computer interaction. However, most users often treat a machine learning model as a black box because of its incomprehensible functions and unclear working mechanism [Fekete, 2013; Liu et al., 2016b; Mhlbacher et al., 2014]. Without a clear understanding of how and why a model works, the development of high-performance models typically relies on a time-consuming trial-and-error process. As a result, academic researchers and industrial practitioners are facing challenges that demand more transparent and explainable systems for better understanding and analyzing machine learning models, especially their inner working mechanisms.

To tackle the aforementioned challenges, there are some initial efforts on interactive model analysis. These efforts have shown that interactive visualization plays a critical role in understanding and analyzing a variety of machine learning models. Recently, DARPA I2O released Explainable Artificial Intelligence (XAI) [DAR, 2016] to encourage research on this topic. The main goal of XAI is to create a suite of machine learning techniques that produce explainable models to enable users to understand, trust, and manage the emerging generation of Artificial Intelligence (AI) systems.

In this paper, we first provide an overview of interactive model analysis. Then we summarize recent interactive model analysis techniques based on their target tasks (such as understanding how a classifier works) [Heimerl et al., 2012]. Research opportunities and future directions are discussed for developing new interactive model analysis techniques and systems.

2. Scope and Overview

We are focused on research and application problems within the context of machine learning. Fig. 1 illustrates a typical machine learning pipeline, from which we first obtain data. Then we extract features that are usable as input to a machine learning model. Next, the model is trained, tested, and gradually refined based on the evaluation results and experience of machine learning experts, a process that is both time consuming and uncertain in building a reliable model. In addition to an explosion of research on better understanding of learning results [Cui et al., 2011, 2014; Dou et al., 2013; Dou & Liu, 2016; Liu et al., 2012, 2014a, 2016c; Wang et al., 2013, 2016], researchers have paid increasing attention to leveraging interactive visualizations to better understand and iteratively improve a machine learning model. The main goal of such research is to reduce human effort when training a reliable and accurate model. We refer to the aforementioned iterative and progressive process as interactive model analysis.

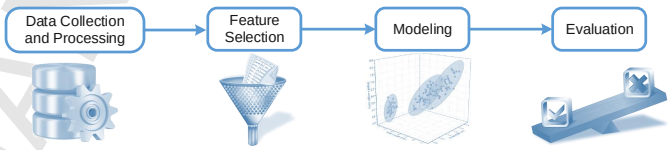

Figure 1: A pipeline of machine learning.

Fig. 2 illustrates the basic idea of interactive model analysis, where machine learning models are seamlessly integrated with state-of-the-art interactive visualization techniques capable of translating models into understandable and useful explanations for an expert. The strategy is to pursue a variety of visual analytics techniques in order to help experts understand, diagnose, and refine a machine learning model. Accordingly, interactive model analysis aims to create a suite of visual analytics techniques that

- understand why machine learning models behave the way they do and why they differ from each other (*understanding*);
- diagnose a training process that fails to converge or does not achieve an acceptable performance (*diagnosis*);
- guide experts to improve the performance and robustness of machine learning models (*refinement*).

*Corresponding author.
Email address: shixia@tsinghua.edu.cn
Received Nov. 28, 2016; Revision accepted Dec. 24, 2016.

DOI: 10.1631/JVI.1000000