Transparent Machine Learning — Revealing Internal States of Machine Learning

Jianlong Zhou
National ICT Australia (NICTA)
13 Garden Street
Eveleigh, NSW 2015 Australia
Jianlong.zhou@nicta.com.au

Yang Wang
National ICT Australia (NICTA)
13 Garden Street
Eveleigh, NSW 2015 Australia
Yang.wang@nicta.com.au

Zhidong Li
National ICT Australia (NICTA)
13 Garden Street
Eveleigh, NSW 2015 Australia
Zhidong.li@nicta.com.au

Fang Chen
National ICT Australia (NICTA)
13 Garden Street
Eveleigh, NSW 2015 Australia
Fang.chen@nicta.com.au

Abstract

This work concerns the revealing internal states of Machine Learning (ML) meaningfully so that users can understand what is going on inside ML and how to accomplish with the learning problem. As a result, ML process becomes more understandable and usable. It changes from a "black-box" to "transparent-box". A case study is presented to show the benefits of transparent ML in improving impact of ML on real-world applications.

IUI 2013, March 19, 2013, Santa Monica, CA, USA.
ACM 978-1-4503-1016-1/12/05.

Keywords

Interactive machine learning; HCI; Black-box; Transparent machine learning.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous. I.2.6. Artificial intelligence: Learning.

Introduction

Machine Learning research is largely inspired by significant problems from various fields such as biology, finance, medicine, society etc.. ML algorithms offer a set of powerful ways to approach those problems that otherwise require manual solution. However, ML research field has a frequent lack of connection between ML research and real-world impact because of complexity of ML methods [1, 2]. For instance, for a domain expert who may not have expertise in ML or programming, an ML algorithm is as a "black-box", where the user defines parameters and input data for the "black-box" and gets output from its running (see Figure 1). This "black-box" approach has obvious drawbacks: it is difficult for users to understand the complicated ML models, such as what is going on inside ML models and how to accomplish with the learning

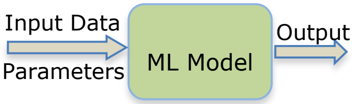


Figure 1. ML model is as a "black-box" to users: users input data and get output from the box. Users do not need to understand what is going on inside or how to accomplish with the problem.

problem. As a result, users are uncertain for ML results and this affects the effectiveness of ML methods.

Our research focuses on making the ML process understandable and usable by end users through revealing internal real-time status update of ML models with meaningful presentations. Because of the leverage of internal states, ML models become transparent to users. The "black-box" ML thus becomes "transparent" ML (TML).

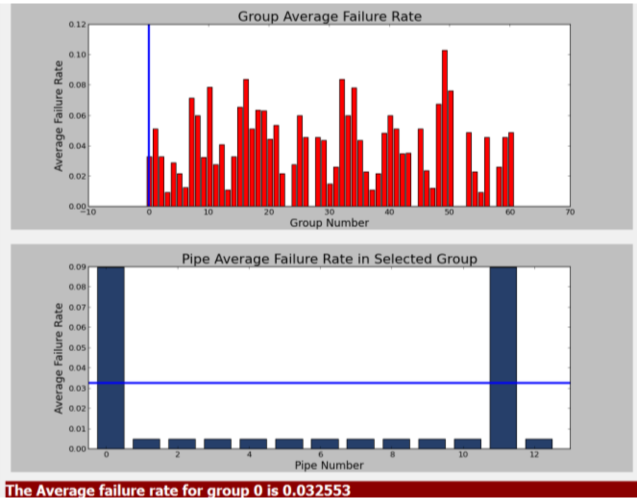
Transparent ML for Interactive Feedback

We propose that interactive ML interfaces must not only supply users with the information on input data and output results, but also enable them to perceive internal real-time status update of ML. As a result, ML process becomes a "transparent-box".

The TML includes following steps:

- Select internal state variables that are dynamically changed and meaningful to users;
- Present the changing of internal state variables visually and meaningfully to users;
- Interact with the ML process (e.g. change ML parameters, insert records) based on transparent feedback from revealing of internal real-time status update.

TML presents selected internal states dynamically to users meaningfully (e.g. money saved, time preserved) with domain knowledge but not only using pure numbers. It provides a feedback loop that aids users learn what is going on and how to accomplish with the given learning problem. Users also have freedom to interact with the ML (e.g. insert data records or add new data features) based on the feedback in order to improve models. TML provides a means for users to assess the model's behaviors against a variety of subjective criteria based on domain knowledge and examples. As a result, the users' understanding and trust of the system could be improved and it benefits the accuracy of learning systems as well. Furthermore, TML allows users progressively improve the learning accuracy by modulating parameters online.

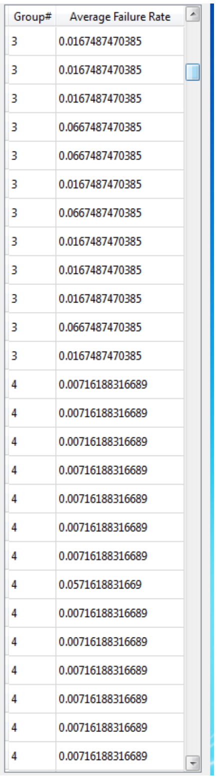


The Average failure rate for group 0 is 0.032553

Figure 2. The real-time status update of ML process is presented to users with interactive graphs and animations.

Case Study

Water supply networks constitute one of the most crucial urban assets. Prediction of water pipe condition through statistical modeling is a significant element for the risk management strategy of water distribution systems. In our previous work, a hierarchical nonparametric model is proposed to predict failure of water pipes. In this model, pipes are divided into *K* groups based on laid years and modeled as a



Group#	Average Failure Rate
3	0.0167487470385
3	0.0167487470385
3	0.0167487470385
3	0.0667487470385
3	0.0667487470385
3	0.0167487470385
3	0.0667487470385
3	0.0167487470385
3	0.0167487470385
3	0.0167487470385
3	0.0667487470385
3	0.0167487470385
4	0.00716188316689
4	0.00716188316689
4	0.00716188316689
4	0.00716188316689
4	0.00716188316689
4	0.00716188316689
4	0.00716188316689
4	0.0571618831669
4	0.00716188316689
4	0.00716188316689
4	0.00716188316689
4	0.00716188316689
4	0.00716188316689
4	0.00716188316689

Figure 3. The real-time status update of ML process is presented to users with a dynamically updated table display.

hierarchical beta process (HBP). In the top level, hyper parameters, that control across all groups of pipes by a beta distribution, are set manually according to domain experts' experience. Then, the mean failure rate (q_k) in each group can be generated from the distribution. In the middle level, the mean failure rate ($pi_{k,i}$) of each pipe asset is generated through another beta distribution with $q_k$as parameter. In the bottom level, the actual failures are generated from a Bernoulli process year by year using $pi_{k,i}$.

In order to allow users better understand how this prediction works, the meaningful internal state of mean failure rate q_k and $pi_{k,i}$ are presented to users interactively. As shown in Figure 2, the top chart presents the status update of q_k and the bottom chart presents the status update of $pi_{k,i}$. During ML process, the charts are dynamically changed to reveal the internal real-time status update. To interact, users can point to any (interact detail) q_k in the top chart and the corresponding $pi_{k,i}$ is presented accordingly in the bottom chart. Furthermore, these real-time status updates can also be accessed through a displayed table as Figure 3. Compared with directly presenting the final prediction of failure rate $\pi_{k,i}$, the presentation of q_k and $pi_{k,i}$ allows users learn how the prediction of failure rate of each pipe is approached. As a result, users' trust on predictions is increased. Therefore, TML benefits the impact of ML on real-world applications.

Contributions of the Work

This work contributed the approach of TML to make ML models understandable and usable by end users.

Acknowledgements

NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program. Authors thank Jinjun Sun, M. Asif Khawaja, and Ronnie Taib for their help in the setup of experiment.

References

[1] Fiebrink, R., and D. Trueman. End-user machine learning in music composition and performance. In *CHI 2012 Workshop on End-User Interactions with Intelligent and Autonomous Systems*. Texas, (2012).

[2] Wagstaff, K. Machine Learning that Matters. In *Proceedings of the 29th International Conference on Machine Learning (ICML)*, Edinburgh, UK, 2012.