ISS-P-066

Click Fraud Prediction with Deep Neural Networks: Challenges and Open Problems

Takashi Sato　　Daniel Berrar

Data Science Laboratory, School of Engineering, Tokyo Institute of Technology

1 Introduction

In the pay-per-click online advertising market, the revenue of a publisher depends on the number of clicks that an advertisement receives. This creates an incentive for dishonest publishers to artificially inflate the click count, for example, by using malware that generates clicks on selected advertisements. This type of cybercrime is known as *click fraud* [2].

Discovering potentially fraudulent activities is a challenging task because clicks generated by sophisticated malware are difficult to distinguish from those that are due to legitimate clicking by humans. However, machine learning methods can detect subtle differences between genuine and machine-generated click traffic.

In this poster presentation, we report challenges and open problems for the prediction of click fraud based on deep neural networks. We analyzed the data sets that were released for the Fraud Detection in Mobile Advertising (FDMA) 2012 Competition [2]. The training and validation sets contain data for 3081 and 3064 publishers, respectively, their advertisements, associated click traffic over three days, and further information, such as country, devices used by the clickers, etc. For each publisher in the training set, a status (OK, Observation, or Fraud) was provided based on a proprietary fraud detection algorithm [2].

2 Challenges and Open Problems

Deep neural networks are excellent methods for representation learning, for example, in image classification tasks. Click log files, however, require data preprocessing based on concrete assumptions about potentially fraudulent patterns or behavior. For example, we may assume that an unusually high number of consecutive clicks from the same IP address in relatively short time intervals is suspicious [1]. Hence, domain knowledge needs to be taken into account for predictive feature engineering.

The second challenge is due to the fact that only a tiny minority of publishers is likely to be fraudulent, leading to highly class-imbalanced data sets with only a few positive examples. For instance, the training and validation set of the FDMA 2012 competition contain only 2.34% and 2.77% publishers with the status Fraud, respectively.

The class imbalance also influences the choice of the performance measure for the evaluation on an independent test set. Publishers should be ranked based on their decreasing likelihood of being fraudulent. The area under the precision-recall curve, called *average precision*, is an adequate measure for this task [2]. Ideally, a machine learning model should use the same performance measure in the training and test phase. The empirical average precision, however, is not continuously differentiable, which hampers its use as a loss function for neural networks.

The major problem in the prediction of click fraud is the lack of a ground truth. The provided class labels in the data sets from the FDMA 2012 competition are based on an in-house developed detection algorithm, which cannot be assumed to work perfectly [1]. For example, several publishers in the training set have the status OK despite clearly suspicious click patterns (such as a high frequency of consecutive clicks in short intervals). Hence, supervised learning is based on inherently uncertain class labels. Obviously, the evaluation and comparison of models is not trivial in the absence of a ground truth.

References

[1] D. Berrar. Learning from automatically labeled data: case study on click fraud prediction. *Knowledge and Information Systems*, 46(2):477–490, 2016.

[2] R. Oentaryo, E.P. Lim, M. Finegold, D. Lo, F. Zhu, C. Phua, E.Y. Cheu, G.E. Yap, K. Sim, M.N. Nguyen, K. Perera, B. Neupane, M. Faisal, Z. Aung, W.L. Woon, W. Chen, D. Patel, and D. Berrar. Detecting click fraud in online advertising: A data mining approach. *Journal of Machine Learning Research*, 15(1):99–140, 2014.