Performance Based Display Advertising: A Large Scale Machine Learning Model in Practice

Krishna Prasad Chitrapura

Yahoo! Labs, Bangalore
pkrishna@yahoo-inc.com

Display advertising is evolving from just selling eye balls to demographic segments, as in traditional print media, to performance based systems where advertisers have an ability to pay for the type of user response valuable to them. Depending on who owns the risk of variance in user response, the advertisers have multiple pricing types options to run their advertising campaign on. These include CPM (pay per view); CPC, (pay per click); CPA, (pay per user action) and Dynamic CPM, which is pay per view but the payment is based on an estimate of the chance of user action. Publishers bear the risk of loss of revenue in the case of CPC and CPA pricing types, whereas the advertisers stand to lose the return in investment in the case of dynamic and CPM pricing types. In addition, the marketplace for display advertising is morphing from being many islands of advertiser and publisher networks to a connected network of networks, which trades over an ad-exchange to better match supply to demand. These changes have brought forth a new set of research challenges such as predicting rates of rare events, managing risk due to variance in prediction and optimal explore-exploit tradeoffs.

In this work, we outlay a subset of new research problems in performance based display advertising and focus on one aspect as a case study of a large scale machine learning model in practice. Specifically, we discuss the anatomy of a response prediction engine that estimates the chances of a page viewer's response to an ad placement. This estimate along with the advertiser's bid is used to allocate an ad to a page that would yield maximal expected revenue. Estimating the chance of rare user events such as clicks and conversion needs state of the art machine learning and robust statistical estimation techniques that can crunch large amount of data periodically on the Grid. There is an inherent trade-off between the complexity of the model and the time taken to re-learn the model periodically. We shall discuss this trade-off in detail and present insights from a live, web-scale advertising exchange.

Further, in the network of networks, we not only face estimation issues due to sparsity of events but also due to *cold starts*, which are advertiser-page view cooccurances that do not appear in the historical data, or appear with very little or no user response. Controlled exploration of such combinations based on the estimate (which exhibits a high variance) is needed to ensure that ads with erroneous estimates do not end up winning large supply of page-views. We will describe one such technique, called *throttling* which mitigates the risk from *cold-starts*.

We also compare and contrast our response prediction techniques to those which are used in popular problems such as movie recommendation[1] [2], collaborative filtering employed by Amazon.com [1] and in paid-search [3].

References

[1] Linden, G.r, Smith, B. and York, J. Amazon.com recommendations: item-to-item collaborative filtering. In *Proc. of IEEE Internet Computing*, Vol. 7, Issues 1. Feb 2003.

[2] Yehuda Koren, Robert Bell and Chris Volinsky. Matrix Factorization Techniques for Recommender Systems. In EEE Computer, 2009.

[3] M. Richardson, E Dominowska and R. Ragno, "Predicting clicks: Estimating the click-through rate for new ads", In *Proceedings of the 16th international conference on World Wide Web (WWW 2007)*, pp 521 - 530, Banff, Canada, 2007.

International Conference on Management of Data
COMAD 2010, Nagpur, India, December 8–10, 2010

[1] http://www.netflixprize.com/