Counterfactual Evaluation and Learning
for
Search, Recommendation and Ad Placement

Thorsten Joachims
Cornell University
Department of Computer Science
Ithaca, NY, USA
tj@cs.cornell.edu

Adith Swaminathan
Cornell University
Department of Computer Science
Ithaca, NY, USA
adith@cs.cornell.edu

ABSTRACT

Online metrics measured through A/B tests have become the gold standard for many evaluation questions. But can we get the same results as A/B tests without actually fielding a new system? And can we train systems to optimize online metrics without subjecting users to an online learning algorithm? This tutorial summarizes and unifies the emerging body of methods on counterfactual evaluation and learning. These counterfactual techniques provide a well-founded way to evaluate and optimize online metrics by exploiting logs of past user interactions. In particular, the tutorial unifies the causal inference, information retrieval, and machine learning view of this problem, providing the basis for future research in this emerging area of great potential impact.
Supplementary material and resources are available online at http://www.cs.cornell.edu/~adith/CfactSIGIR2016.

CCS Concepts

•Information systems → Information retrieval; Retrieval models and ranking; Evaluation of retrieval results;

Keywords

Learning to Rank; Counterfactual Estimation; Causal Inference; Batch Learning from Bandit Feedback

Motivation

How many clicks will a new ad placement system get? Will a different news-ranking algorithm increase the dwell times of my users? What ranking function will minimize abandonment? These and similar questions are crucial to consider when building good search engines, recommendation systems, ad placement systems and most other systems that interact with users.

One approach to address these questions is to field the new system in an A/B test. Unfortunately, this process is slow, costly and puts tight limits on the number of systems that can be fielded. However, we typically have large amounts of log-data from previous versions of our systems. Can we use this old data to evaluate new systems?

Trying to answer operational questions about a new system B with log-data from an old system A is essentially a counterfactual problem: how would system B have performed on some user-centered metric, if it had been operational instead of system A? (e.g. how many clicks would the new ad placement system have gotten, if it had been deployed in the past?) While it may seem like the logs from system A will be insufficient for estimating the properties of system B, research over the past years [21, 3, 7, 13, 23] has shown the opposite: the logs of system A can provide unbiased counterfactual estimates for the performance of system B, even though B has never been fielded.

This opens up a wide area of research on new evaluation and learning techniques for information retrieval. Unlike deploying new systems in weeks-long A/B tests to get unbiased performance estimates [10], counterfactual estimators can provide unbiased estimates from existing logs without delay. Analogously, counterfactual learning techniques can use existing log files for learning improved systems, without any need for online learning methods that require interactive control. In this way, this new approach to evaluation and learning can revolutionize the process of developing and optimizing systems that interact with users.

Research on counterfactual estimation (also called estimation of treatment effects and off-policy evaluation) and counterfactual learning (also called policy optimization) has made rapid progress in recent years [8, 3, 7, 6, 2, 13, 22, 9, 11, 14, 15, 19, 24, 12, 18], and it is starting to have impact in commercial applications. The recent workshop at WWW2015 on Online and Offline Evaluation of Web-based Services [8] highlights keen interest in these techniques from academia and industry alike. However, this research has happened in many rather disjoint fields, including statistics [16], economics [1], reinforcement learning [20], contextual bandit learning [21], Monte Carlo estimation [17] and information retrieval [23, 5, 4]. It is an excellent time to aggregate and unify the existing works into one coherent tutorial that is broadly accessible to an information retrieval audience.

Objectives

This tutorial will provide an overview of the rapidly growing body of research on offline evaluation and learning using online metrics. A confluence of developments in machine learning, causal inference, economics, and information retrieval has recently pushed these methods to be well-founded and practical. This tutorial will provide a unifying view that will enable newcomers to enter the field, enabling broader research and further adoption of these methods in practice. The objectives of the tutorial are:

1. Provide a unified view of counterfactual estimation and learning techniques as they apply to Information Retrieval evaluation and learning.

2. Translate the disparate terminology from economics, causal inference, and machine learning into a framework accessible to Information Retrieval researchers.

3. Give an overview of counterfactual evaluation techniques and how they apply to IR problems.

4. Give an overview of counterfactual learning techniques and how they apply to IR problems.

5. Demonstrate strengths and limitations of counterfactual techniques via case studies on IR problems.

6. Provide publicly available benchmark datasets to enable IR research in this area.

7. Outline directions for future research.

Prerequisites

This tutorial is aimed at an audience with intermediate experience with information retrieval. It assumes the following prerequisites:

- Familiarity with standard IR methods, applications and evaluation metrics is assumed and only briefly reviewed.

- Basic understanding of probability theory and introductory statistics is sufficient for understanding most of the tutorial.

- Some topics require basic understanding of machine learning.

All code samples demonstrating counterfactual analysis for IR will be in Python3. Participants who wish to run these demos locally must bring a device capable of running Python3 scripts. Supplementary material for this tutorial is available online at http://www.cs.cornell.edu/~adith/CfactSIGIR2016.

Presenters

Thorsten Joachims is a Professor in the Department of Computer Science and in the Department of Information Science at Cornell University. His research interests center on a synthesis of theory and system building in machine learning, with applications in information retrieval and recommendation. His past research focused on support vector machines, learning to rank, learning with preferences, and learning from implicit feedback, text classification and structured output prediction. He is an ACM Fellow, AAAI Fellow and Humboldt Fellow.

Adith Swaminathan is a PhD candidate in the Department of Computer Science at Cornell University, advised by Prof. Thorsten Joachims. His research interests are at the core of this tutorial, focusing on principles and algorithms for off-policy evaluation and learning for retrieval and recommendation systems. He received a BTech degree in Computer Science and Engineering from IIT Bombay in 2010 and MSc in Computer Science from Cornell University in 2014.

Acknowledgements

1. REFERENCES

[1] S. Athey and G. Imbens. Recursive Partitioning for Heterogeneous Causal Effects. *ArXiv e-prints*, 2015.

[2] A. Beygelzimer and J. Langford. The offset tree for learning with partial labels. In *KDD*, pages 129–138, 2009.

[3] L. Bottou, J. Peters, J. Q. Candela, D. X. Charles, M. Chickering, E. Portugaly, D. Ray, P. Y. Simard, and E. Snelson. Counterfactual reasoning and learning systems: The example of computational advertising. *Journal of Machine Learning Research*, 14(1):3207–3260, 2013.

[4] B. Carterette, E. Kanoulas, V. Pavlu, and H. Fang. Reusable test collections through experimental design. In *SIGIR*, pages 547–554, 2010.

[5] B. Carterette, E. Kanoulas, and E. Yilmaz. Advances on the development of evaluation measures. In *SIGIR*, pages 1200–1201, 2012.

[6] M. Dudík, D. Erhan, J. Langford, and L. Li. Doubly robust policy evaluation and optimization. *Statistical Science*, pages 485–511, 2014.

[7] M. Dudík, J. Langford, and L. Li. Doubly robust policy evaluation and learning. In *ICML*, pages 1097–1104, 2011.

[8] N. Gupta, E. Koh, and L. Li. Workshop on online and offline evaluation of web-based services. In *WWW Companion*, 2015.

[9] K. Hofmann, A. Schuth, S. Whiteson, and M. de Rijke. Reusing historical interaction data for faster online learning to rank for IR. In *WSDM*, pages 183–192, 2013.

[10] R. Kohavi, R. Longbotham, D. Sommerfield, and R. M. Henne. Controlled experiments on the web: survey and practical guide. *Data Mining and Knowledge Discovery*, pages 140–181, 2009.

[11] J. Langford, A. Strehl, and J. Wortman. Exploration scavenging. In *ICML*, pages 528–535, 2008.

[12] L. Li, S. Chen, J. Kleban, and A. Gupta. Counterfactual estimation and optimization of click metrics in search engines: A case study. In *WWW Companion*, pages 929–934, 2015.

[13] L. Li, W. Chu, J. Langford, and X. Wang. Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In *WSDM*, pages 297–306, 2011.

[14] L. Li, R. Munos, and C. Szepesvari. Toward minimax off-policy value estimation. In *AISTATS*, 2015.

[15] J. Mary, P. Preux, and O. Nicol. Improving offline evaluation of contextual bandit algorithms via bootstrapping techniques. In *ICML*, pages 172–180, 2014.

[16] P. Rosenbaum and D. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[17] R. Rubinstein and D. Kroese. *Simulation and the Monte Carlo Method*. Wiley, 2008.

[18] T. Schnabel, A. Swaminathan, A. Singh, N. Chandak, and T. Joachims. Recommendations as treatments: Debiasing learning and evaluation. *ArXiv e-prints*, 2016.

[19] A. L. Strehl, J. Langford, L. Li, and S. Kakade. Learning from logged implicit exploration data. In *NIPS*, pages 2217–2225, 2010.

[20] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, 1998.

[21] A. Swaminathan and T. Joachims. Counterfactual risk minimization: Learning from logged bandit feedback. In *ICML*, pages 814–823, 2015.

[22] A. Swaminathan and T. Joachims. The self-normalized estimator for counterfactual learning. In *NIPS*, pages 3213–3221, 2015.

[23] E. Yilmaz, E. Kanoulas, and J. A. Aslam. A simple and efficient sampling method for estimating AP and NDCG. In *SIGIR*, pages 603–610, 2008.

[24] B. Zadrozny, J. Langford, and N. Abe. Cost-sensitive learning by cost-proportionate example weighting. In *ICDM*, pages 435–, 2003.