From Offline to Online Experimentation: Considerations from Experiences at Spotify

Ben Carterette

Spotify

ABSTRACT

The history of experimenting on information access systems using offline test collections—the Cranfield paradigm—goes back many decades and is a major aspect of scientific progress in search and IA. Its wide-scale adoption has been driven in part by its robustness and ease of use, in part by evaluation workshops like NTCIR, and in part by the emergence of new information access scenarios and problems that can adapt it. Despite that, there is a lot we still don't know about the ability of offline experiments to predict online outcomes with real users in real-world conditions. In this talk I discuss a common framework for thinking about experimentation and connect it to both offline Cranfield experiments and online A/B testing. Using examples from Spotify search and recommendation, I show how offline experiments motivate online development and vice versa. Developing a better understanding of how offline experiments translate into online experiences will be key as approaches from our research continue to be adopted into real-world technology.

BIOGRAPHY

Ben Carterette is a Senior Research Manager at Spotify, where he leads a team of research scientists investigating problems such as ML for search and recommendation, offline and online experimentation, user models of consumption and satisfaction, and models of music and podcast content. He was formerly an Associate Professor at the University of Delaware, where he maintains an affiliated position. He has published over a hundred papers in information retrieval and data mining venues and been co-author of five Best Paper Award-winning works. With collaborators, he has co-organized 12 TREC tracks on various topics, and has served as a standing member of the TREC PC and as an NTCIR PC member for several years. Dr. Carterette is currently serving as Chair of the ACM SIGIR Executive Committee.