

Using Machine Learning to Improve the Email Experience

Marc Najork
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA, USA
najork@google.com

ABSTRACT

Email is an essential communication medium for billions of people, with most users relying on web-based email services. Two recent trends are changing the email experience: smartphones have become the primary tool for accessing online services including email, and machine learning has come of age. Smartphones have a number of compelling properties (they are location-aware, usually with us, and allow us to record and share photos and videos), but they also have a few limitations, notably limited screen size and small and tedious virtual keyboards. Over the past few years, Google researchers and engineers have leveraged machine learning to ameliorate these weaknesses, and in the process created novel experiences. In this talk, I will give three examples of machine learning improving the email experience.

The first example describes how we are improving email search. Displaying the most relevant results as the query is being typed is particularly useful on smartphones due to the aforementioned limitations. Combining hand-crafted and machine-learned rankers is powerful, but training learned rankers requires a relevance-labeled training set. User privacy prohibits us from employing raters to produce relevance labels. Instead, we leverage implicit feedback (namely clicks) provided by the users themselves. Using click logs as training data in a learning-to-rank setting is intriguing, since there is a vast and continuous supply of fresh training data. However, the click stream is biased towards queries that receive more clicks – e.g. queries for which we already return the best result in the top-ranked position. I will summarize our work [2] on neutralizing that bias.

The second example describes how we extract key information from appointment and reservation emails and surface it at the appropriate time as a reminder on the user's smartphone. Our basic approach [3] is to learn the templates that were used to generate these emails, use these templates to extract key information such as places, dates and times, store the extracted records in a personal information store, and surface them at the right time, taking contextual information such as estimated transit time into account.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CIKM'16 October 24-28, 2016, Indianapolis, IN, USA

© 2016 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4073-1/16/10.

DOI: <http://dx.doi.org/10.1145/2983323.2983371>

The third example describes Smart Reply [1], a system that offers a set of three short responses to those incoming emails for which a short response is appropriate, allowing users to respond quickly with just a few taps, without typing or involving voice-to-text transcription. The basic approach is to learn a model of likely short responses to original emails from the corpus, and then to apply the model whenever a new message arrives. Other considerations include offering a set of responses that are all appropriate and yet diverse, and triggering only when sufficiently confident that each response is of high quality and appropriate.

CCS Concepts

•Information systems → Email; •Computing methodologies → Machine learning;

Keywords

Email; Information Extraction; Machine Learning; Ranking

Bio

Marc Najork is a Senior Staff Research Scientist at Google, where he manages a team working on a portfolio of machine learning problems. Before joining Google in 2014, Marc spent 12 years at Microsoft Research Silicon Valley and 8 years at Digital Equipment Corporation's Systems Research Center in Palo Alto. Much of his past research has focused on improving web search, and on understanding the evolving nature of the web. Marc has published about 60 papers and holds 25 issued patents. He received a Ph.D. in Computer Science from the University of Illinois at Urbana-Champaign.



References

- [1] A. Kannan, K. Kurach, S. Ravi, T. Kaufmann, A. Tomkins, B. Miklos, G. Corrado, L. Lukács, M. Ganea, P. Young, and V. Ramavajjala. Smart Reply: Automated response suggestion for email. In *22nd International Conference on Knowledge Discovery and Data Mining (KDD)*, 2016.
- [2] X. Wang, M. Bendersky, D. Metzler, and M. Najork. Learning to rank with selection bias in personal search. In *39th International Conference on Research and Development in Information Retrieval (SIGIR)*, 2016.
- [3] W. Zhang, A. Ahmed, J. Yang, V. Josifovski, and A. J. Smola. Annotating needles in the haystack without looking: Product information extraction from emails. In *21st International Conference on Knowledge Discovery and Data Mining (KDD)*, 2015.