

Item-to-item Recommendations at Pinterest

Stephanie Rogers
Pinterest
San Francisco, California
srogers@pinterest.com

ABSTRACT

This talk presents Pinterest Related Pins, an item-to-item recommendation system that combines collaborative filtering with content-based ranking to drive a quarter of the total engagement on Pinterest. Signals derived from user curation, the activity of users organizing content, are highly effective when used in conjunction with content based ranking. This will be an in-depth dive into the end-to-end system of Related Pins, a real-world implementation of an item-to-item hybrid recommendation system.

Candidate Generation

Pinterest is a user-powered content curation service as content is collected and grouped into boards, creating a rich set of metadata associated with Pins' images. User curation reveals relations among images: we observed that images of Pins on the same board are semantically and often visually similar to each other. Therefore, if enough users save these two Pins together, there is a high likelihood that another user may also find them to be related.

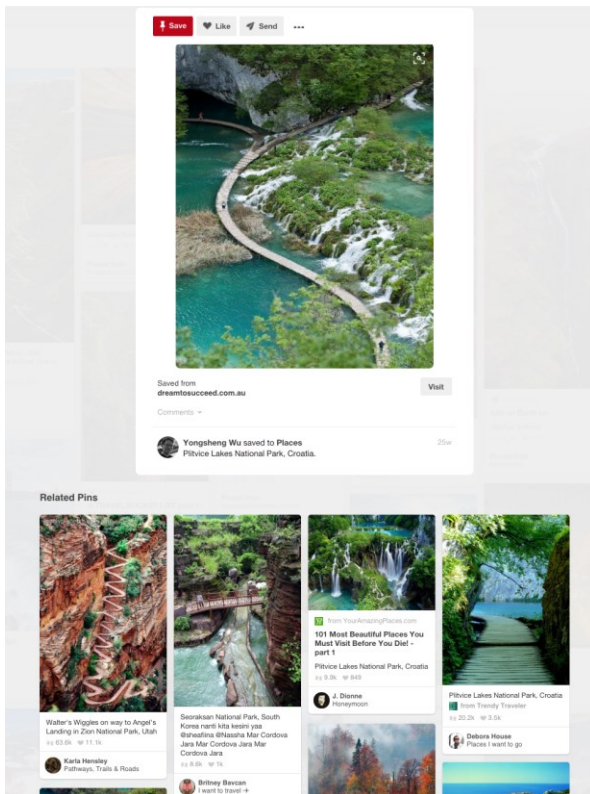
Using board co-occurrence as an estimate of how similar two pins are, we can narrow down the set of pins we believe to be related. This form of collaborative filtering allows us to generate an initial, smaller set of image candidates, making it possible and more feasible to do content-based reranking.

Reranking

After generating a set of candidates for each pin, we rerank the results using a machine learned model based on various features. These features are a combination of pin features, both from the query and candidate pins, and user context features to personalize the results. We have explored various modeling techniques from linear modeling to gradient boosted decision trees.

Challenges

Along the way we have been presented with a ton of challenges including: personalization, the tradeoff of relevance vs. diversity, infrastructure stability and pipeline productionization, collecting unbiased training data, and more. Many of these challenges will be presented in the talk.



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.
Copyright is held by the owner/author(s).
RecSys '16, September 15-19, 2016, Boston, MA, USA
ACM 978-1-4503-4035-9/16/09.
<http://dx.doi.org/10.1145/2959100.2959130>