

Spinning the AI Pinwheel

Jai Ranganathan
KeepTruckin
jai@keeptruckin.com

ABSTRACT

Advances in supervised machine learning have frequently been fueled by access to large-scale labeled data. In business, however, natural labels may not exist. In these cases, a common industry playbook involves using manual, human annotation to label enough data points in order to train a model. This paper describes a general structure for accelerating the annotation process using artificial intelligence and combining it with model quality assurance (QA).

In this talk we walk through this process in detail. We start with rich manual annotation of a small number of unlabeled data points. These can then be used to train a series of coarse predictive models that are used to prepopulate some default selections in the annotation tool and speed up annotator performance. With more data points, models can be retrained on a regular cadence and less human intervention is required. Finally, models can provide defaults for all fields, and re-training continues until the annotator override rate reaches a production-grade level.

Tradeoffs of this type of approach include balancing the increased annotation efficiency with engineering costs associated with building annotation and quality assurance tools. We will walk through these tradeoffs, which depend on the problem class and complexity of the model.

Finally, we will include a detailed industry case study based on the use of artificial intelligence in the annotation process at KeepTruckin, where we use annotation to label vehicle location history data.

BIOGRAPHY

Jai Ranganathan is VP of Product & Data Science at KeepTruckin. Prior to joining KeepTruckin, Jai worked at Uber where he served as Senior Director of Data Science & Product, managing Machine Learning & AI, data, marketing systems, and operations tooling. Before that, Jai served as Senior Director of Product leading Machine Learning at Cloudera.



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.

KDD '19, August 4–8, 2019, Anchorage, AK, USA.

© 2019 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-6201-6/19/08.

DOI: <https://doi.org/10.1145/3292500.3340405>