DATA SCIENCE

How does TikTok's algorithm know me so well?

A high level overview of how personalised feed algorithms work

Kaushik Sureshkumar

Dec 19, 2020 5 min read

Photo by Hello I’m Nik ð on Unsplash

The phrase "knows me better than I know myself" was mentioned a few times when discussing TikTok's algorithm with friends. This is a mark of a great recommendation system, so I wanted to look at what makes it so good from a data point of view.

The data science team at TikTok have no doubt worked on many different recommendation systems involving complex machine learning techniques. This post, however, will focus on how the data that TikTok uses for this exercise, along with the format of content on the platform, already give it a massive advantage.

To give you a better understand of why this is, let me briefly introduce how recommendation systems work. These algorithms are built on what's called a **user-item matrix**. This is essentially a table where each row represents an individual user and each column represents an individual item that the algorithm needs to recommend. For example, at Spotify these items would be songs, at Netflix they'd be films and at Tinder they'd be other users.

The value in each cell of the table represents some sort of interaction between the user and the item. This could be whether a user liked a post (Instagram), a star rating of a product after use (Amazon) or whether they finished listening to a song (Spotify). We call this **feedback** since it gives us information on whether the user liked the item or not and, in some cases, how much they liked it.

Using this table, given a user, we can work out other similar users, and recommend items those users have watched/bought etc. Given items we can recommend other items based on what other users have bought with this item etc.

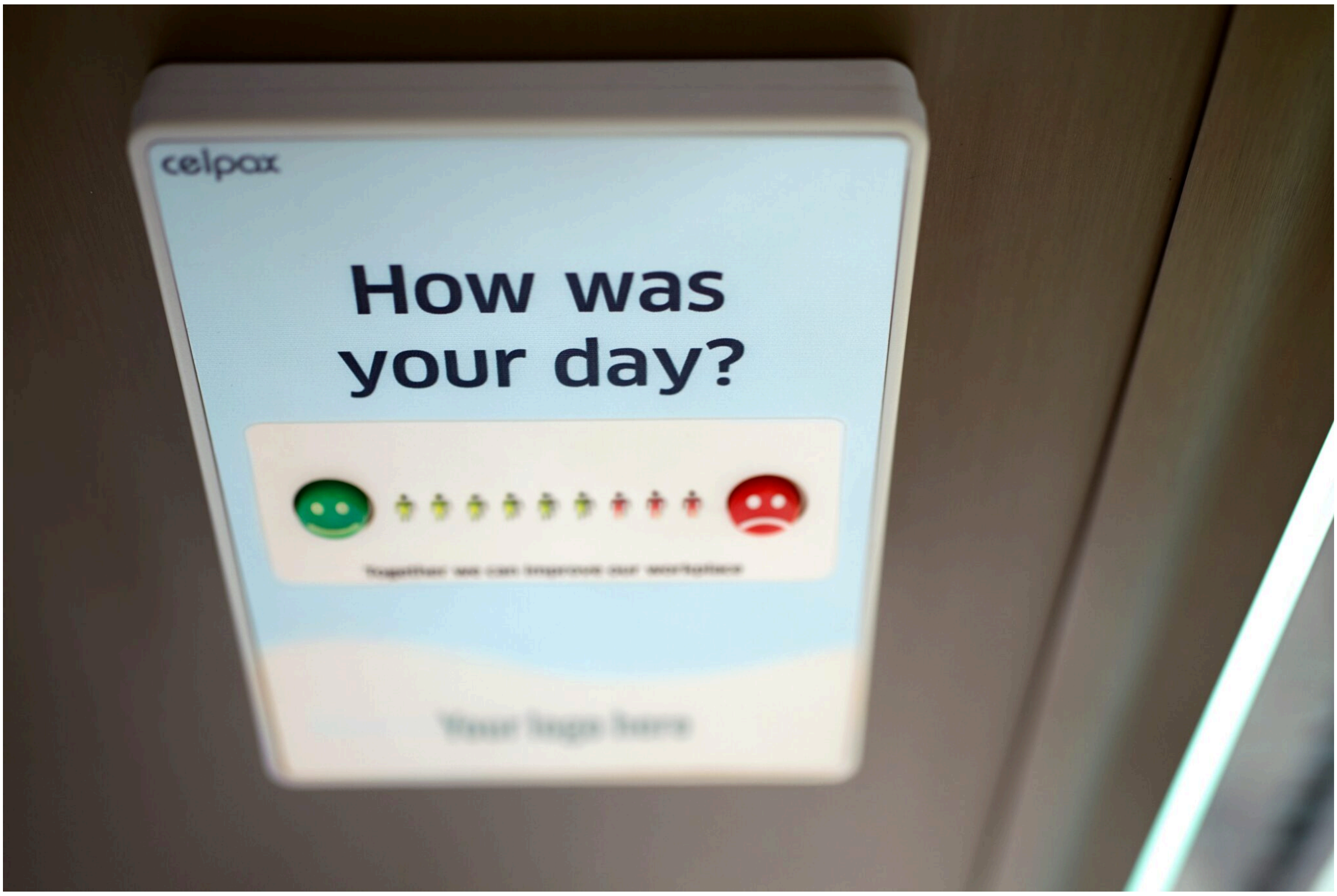


Photo by Celpax on Unsplash

How does this relate to quality of recommendations?

The quality of feedback has a big impact on the quality of recommendations. Higher quality information on the user's preferences leads to higher quality recommendations. In particular, there are two main factors that affect the quality of feedback, implicit vs explicit and the granularity of the feedback.

Implicit feedback is what we call subconscious and natural user interactions that we can use as a proxy for the user's true preferences. Whether they reacted to a post (LinkedIn), whether they clicked on a playlist (Spotify) and whether they followed another user (Twitter) are some examples of implicit feedback.

Explicit feedback refers to actual ratings that users have given to items. This conveys more information than implicit feedback since we don't need to guess what the user's true preferences are. User ratings of products (Amazon), films (Netflix) and apps (App Store) are some examples of explicit feedback. Although explicit feedback gives us more information, it's much harder to acquire this data. Since rating items requires more effort from the user, fewer of them will actually do it. The result is a very sparse user-item matrix with few entries, leading to a less than optimal recommendation algorithm.

Granularity of the feedback refers to how much information we get from it. A rating scale out of 10 gives us more information than a rating scale out of 5 which in turn gives us more information than a rating scale out of 2. It allows us to understand not just whether a user liked or disliked an item, but also how much they liked or disliked it.

Most implicit feedback data is binary – did a user perform a specific action on this item or not – and therefore less granular than explicit feedback data, however it's much easier to acquire as it requires less effort from the user.

Photo by Joey Huang on Unsplash

TikTok’s main advantages

TikTok is able to use the time a user spends on each video as a form of **granular implicit feedback**. Not only is it granular but since it’s a time-based scale it is continuous, so it’s a very accurate proxy for the user’s true preferences. Since it doesn’t require any extra effort from the user in order to gather this data, TikTok is able to do this successfully at scale, leading to super accurate recommendations.

This kind of granular implicit feedback is something that Netflix and Spotify could also use, and no doubt they do. However this is where TikTok’s second major advantage makes a difference – the format of content on their platform. Short form videos are much easier to consume than films, episodes and even songs. In a given TikTok session, a user is likely to scroll through 20 different videos, creating 20 new entries in the user-item matrix. Whereas in a Netflix session, a user is likely to only watch 1 film or TV show, creating just 1 new entry in the user-item matrix. This allows TikTok to gather this data at scale very easily, resulting in a very accurate algorithm.

How does Machine Learning fit into this?

Building, storing and computing on this user-item matrix is extremely computationally expensive. So we use ML methods like Matrix Factorisation or neural net methods like Deep Collaborative Filtering to learn user and item representations (matrices or embeddings, respectively) which are easier to store and compute on.

Data Scientists at TikTok use these Machine Learning and Deep Learning techniques on the user-item matrix to create great recommendation algorithms. However, granular implicit feedback, along with the fact that the product is built around short form videos, the easiest form of media to consume, is what allows them to gather this vast wealth of data and use it so efficiently to perfect their feed algorithms.

Thanks for reading this article! I hope it helped you get a better understanding of how TikTok is able to serve very accurate recommendations.

If you enjoy reading my articles, would like to support my writing and are thinking of getting a Medium subscription, feel free to use my referral link below. I'd get a percentage of your subscription fees.

Join Medium with my referral link – Kaushik Sureshkumar

• • •

WRITTEN BY

Kaushik Sureshkumar

See all from Kaushik Sureshkumar

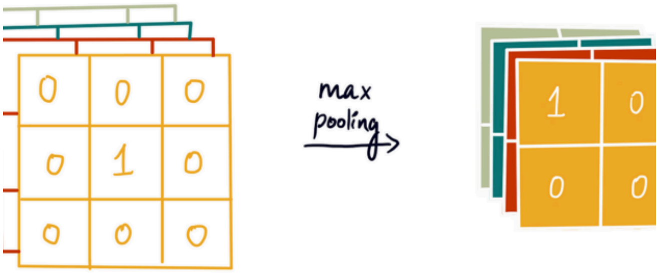
Topics:

Data Science Product Recommendation System

Recommendations Social Media

Share this article:

Related Articles



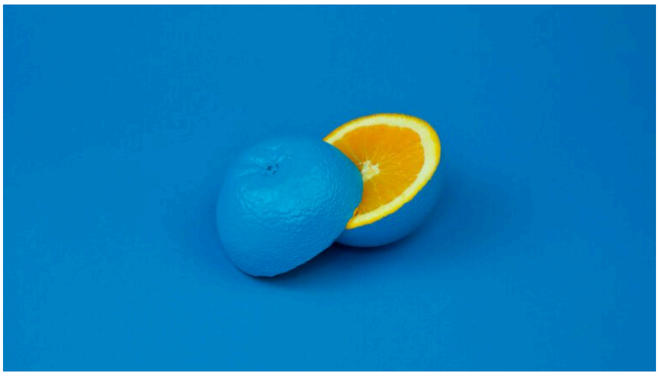
ARTIFICIAL INTELLIGENCE

Implementing Convolutional Neural Networks in TensorFlow

Step-by-step code guide to building a Convolutional Neural Network

Shreya Rao

August 20, 2024 6 min read



DATA SCIENCE

Hands-on Time Series Anomaly Detection using Autoencoders, with Python

Here's how to use Autoencoders to detect signals with anomalies in a few lines of...

Piero Paialunga

August 21, 2024 12 min read

DATA SCIENCE

Solving a Constrained Project Scheduling Problem with Quantum Annealing

Solving the resource constrained project scheduling problem (RCPSP) with D-Wave's hybrid constrained quadratic model (CQM)

Luis Fernando PÉREZ ARMAS, Ph.D.

August 20, 2024 29 min read

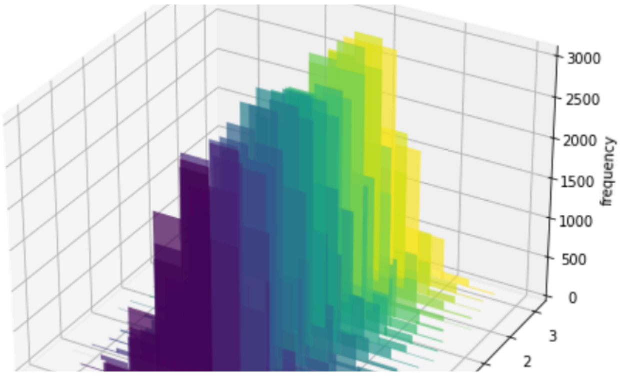
DATA SCIENCE

Back To Basics, Part Uno: Linear Regression and Cost Function

An illustrated guide on essential machine learning concepts

Shreya Rao

February 3, 2023 6 min read



DATA SCIENCE

Must-Know in Statistics: The Bivariate Normal Projection Explained

Derivation and practical examples of this powerful concept

Luigi Battistoni

August 14, 2024 7 min read

DATA SCIENCE

Our Columns

Columns on TDS are carefully curated collections of posts on a particular idea or category…

TDS Editors

November 14, 2020 4 min read

DATA SCIENCE

Optimizing Marketing Campaigns with Budgeted Multi-Armed Bandits

With demos, our new solution, and a video

Vadim Arzamasov

August 16, 2024 10 min read

ABOUT · ADVERTISE · PRIVACY POLICY · TERMS OF USE

COOKIES SETTINGS