

Playlist Optimisation as a Traveling Salesperson Problem



Grant Holtes · [Follow](#)

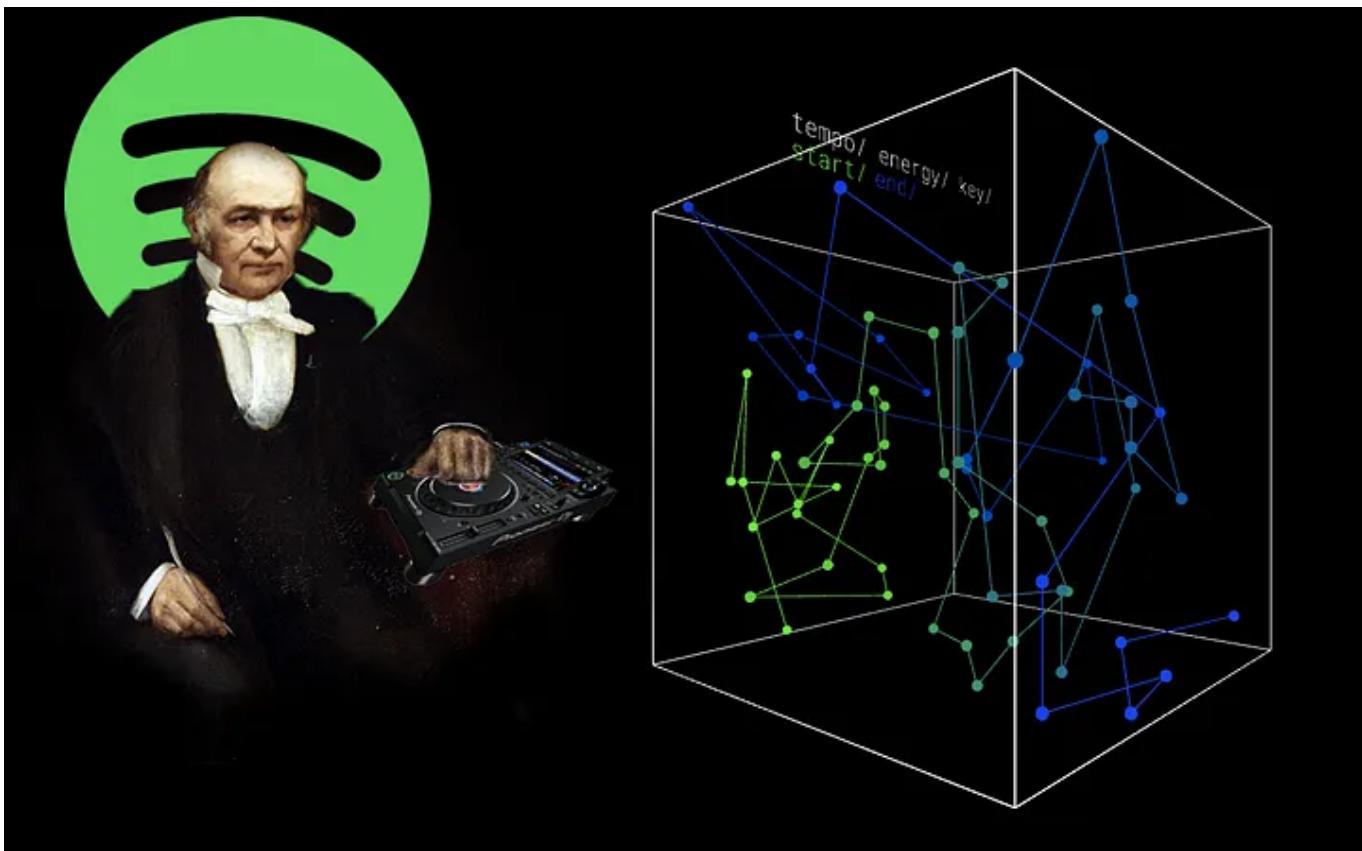
Published in DataSeries · 5 min read · Dec 25, 2021

12



...

For somebody who takes a lot of pride in their playlists, mine are pretty sub-par. Sure, I may like the songs, but the order of playback leaves a lot to be desired — Pitbull and The Veronicas certainly belong in the same playlist, but you might not want their top hits back-to-back! This is my over-engineered solution to this minor issue.



Motivation

Spotify is a fantastic service which has drastically changed how we find, listen and share music. Their tech forward approach has also allowed new features to be released that further elevate their user's ability to interact with their music, such as the *discover weekly* playlist or the recently released blend tool that merges two people's music tastes into a single playlist.

At the heart of the Spotify user experience is the playlist, which allows the user a way to catalogue and organise songs in a manner reminiscent of the cassette mixtapes of the 1980s. Without playlists the listening experience on Spotify would be a transient one, with users needing to search through the endless library of tracks to find “their” music. As such playlists provide the user with some level of ownership and control over what is “their” music in a system with no ownership and a universe of music and genres.

Given the importance of playlists and the mechanism by which new songs are added to a playlist (just slap them on the end), it is unsurprising that personal playlists can quickly become incoherent, a consequence of evolving tastes and the level of effort required to reorder and reorganise songs.

To avoid this fate for my own playlists, I have started to build solutions to this issue through using Spotify's developer tools to create my own methods to interact with the Spotify music universe.

This is part two in a series, with links to the other parts at the bottom of this article.

Playlist Optimisation

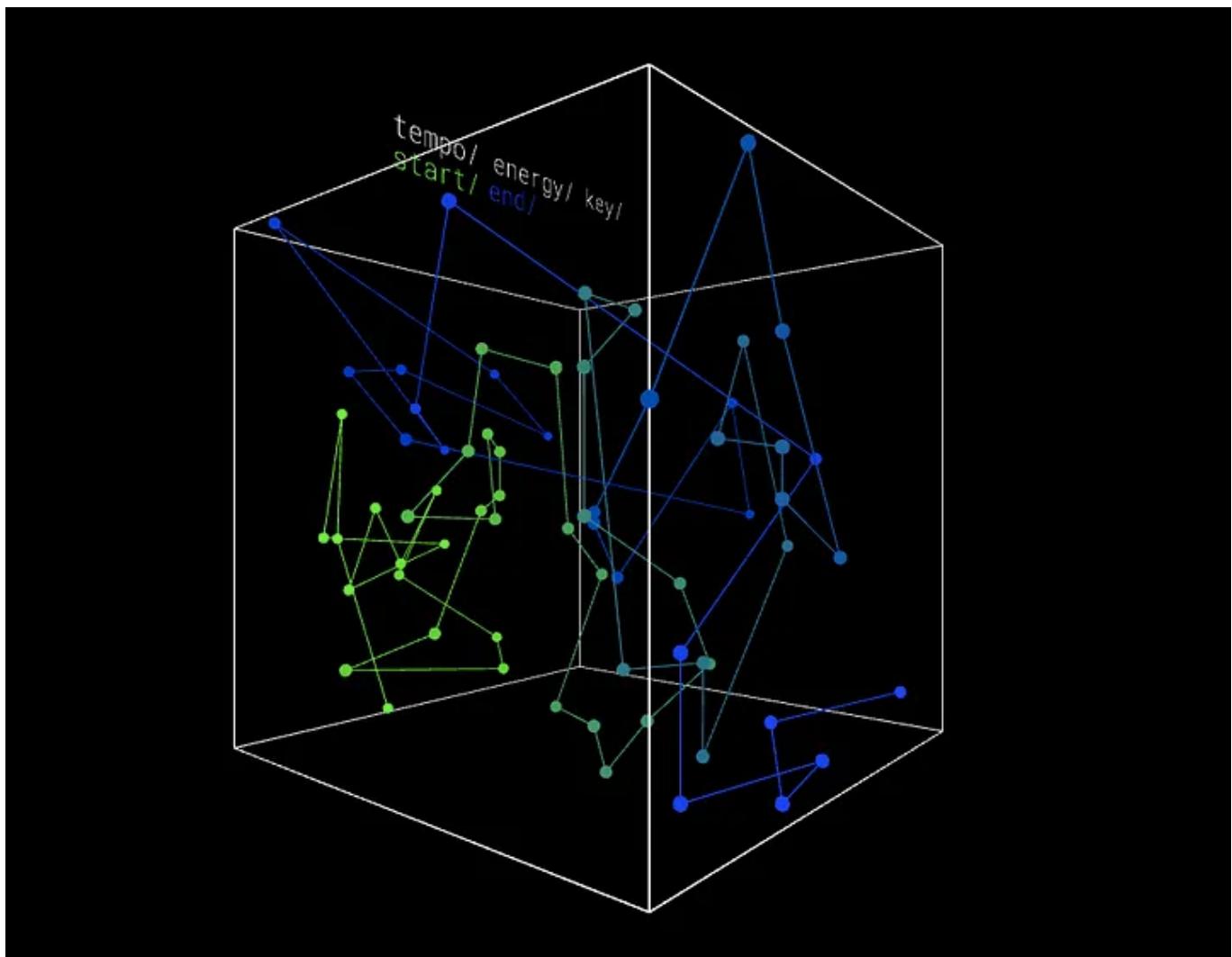
“DJing is a complex analytical as well as creative task, creating an automatic DJ has proved to be highly nontrivial” Len Vande Veire and Tijl De Bie, EURASIP Journal on Audio, Speech, and Music Processing (2018)

User's Spotify playlists tend to be chronologically organised by the date when a song is added, which does not typically lead to a well ordered playback. When seeking to improve this order, it helps to reframe the problem as “what order should we play the songs such that every subsequent song is acceptably similar to the previous song”.

One shortcut to determine how similar two songs is to use the Spotify audio analysis data, which includes measurements of tempo, key, as well as more subjective measures such as “danceability” and “acousticness” which capture the sound and mood of the song. If a set of these measures are stored as vectors for each song, the similarity between songs can be quantified as the distance between these vectors, with various dimensions weighted by relative importance. Two songs that are nearby as measure by their vectors

should have a similar BPM, key and general mood, while songs that are further apart may be very different and so unsuitable to play back-to-back.

The task of ordering the songs of the playlist to minimise the sum of the distance between sequential songs is version of the traveling salesman problem or TSP, which seeks find a route between cities that minimise the total distance traveled. The only major difference is that instead of a two dimensional real-world space, we have an arbitrary number of dimensions in the song vectors.

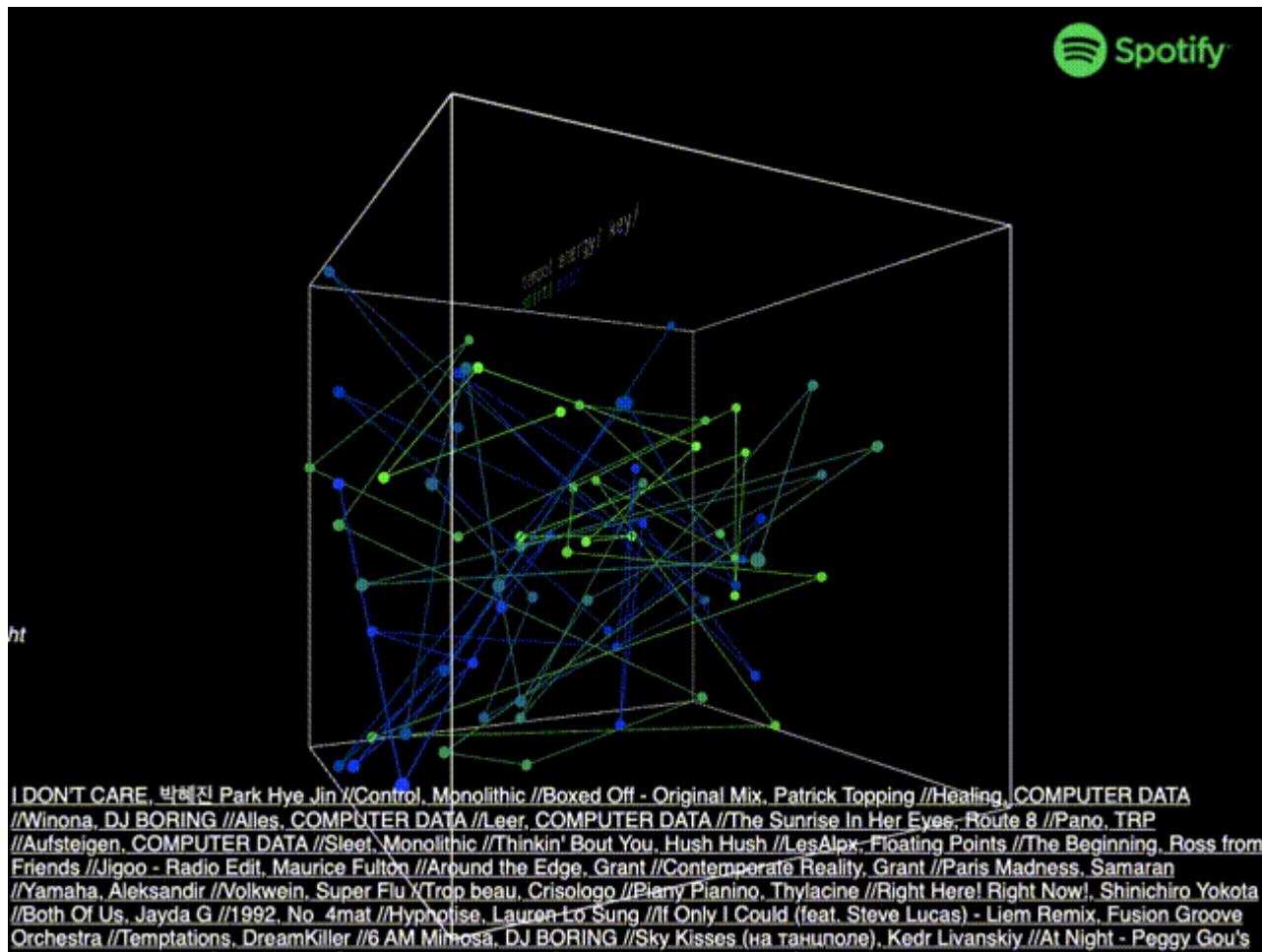


Visualisation of a TSP solution, showing a subset of the song vector dimensions: Tempo, key and “energy”

Finding an exact solution to the TSP is a famously NP-hard problem, which requires an infeasible amount of computational power for larger numbers of data points. The complexity of the TSP is $O(n!)$, which becomes impractical beyond roughly 20 data points, and even optimised versions are impractical beyond 50 data points. As we don't know how long a user's playlist may be and want the algorithm to run in a standard web browser, seeking an exact solution is unacceptable.

Instead an approximate solution is used, in this case a simple nearest neighbours approach is used which has a complexity of $O(n$ squared). The algorithm can be summarised as the following steps.

1. Mark all points as unseen.
2. Select a point from the set, in this case the array of playlist vectors.
Append the point to the output array and mark it as seen.
3. While there are unseen points, find the nearest point by looping over all unseen points and measuring the distance to the previously seen point.
4. Append the nearest point to the output array and mark it as seen.
5. Return the output array



The solution visualiser for a unoptimised playlist, followed by a number of candidate solutions

This greedy approach is far from optimal, but is significantly faster and simpler than competing approaches, and is well suited to this application where speed is preferred over an exact solution. An additional perk of this method is that every starting point yields a slightly different approximate solution, which allows the user to shuffle between approximately optimal solutions by simply recomputing the solution, given that the starting point is randomly chosen. When this is combined with a visualisation of the solution, it allows the user to pick the solution they like the look of best. This solution could be further improved by selecting a starting point with a desirable characteristics for the start of a playlist, such as low BPM compared to the other songs.

The tool

For those that are interested the playlist optimiser tool is available at
<https://www.grantholtes.com/smartschedule>

Other stories

- Part 1: [Building Better Playlists: Spotify Integration](#)

Algorithms

Machine Learning

Spotify

Web Development

Music



Written by Grant Holtes

242 Followers · Writer for DataSeries

AI and Analytics Consultant in Melbourne. www.grantholtes.com

Follow



More from Grant Holtes and DataSeries



 Grant Holtes in Towards Data Science

How Animal Investors Beat the Market

3 min read · Apr 14, 2018

 124 

 Muhammad Danyal in DataSeries

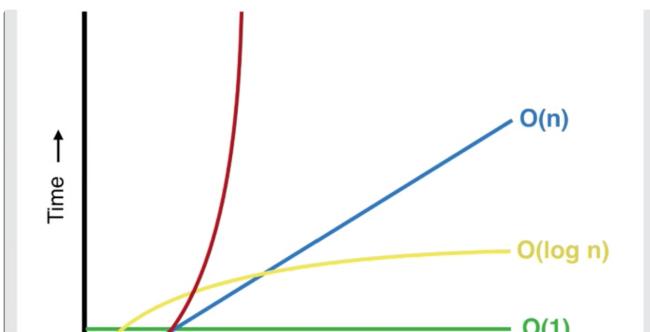
Public Claims and How to validate a JWT

JWT stand for JSON Web Token. It is a security validation mechanism widely used...

7 min read · Apr 14, 2020

 475 



 Maxwell Harvey Croy in DataSeries

How To Calculate Time Complexity With Big O Notation

Part 2 of a series breaking down Big O Notation and Time and Space Complexity fo...

 · 6 min read · Jul 28, 2020

 763 

 Grant Holtes in Towards Data Science

Comparison of Dimensionality Reduction Techniques for Image...

This task was undertaken as part of a proof of concept for Look and Learn, an online library...

4 min read · Dec 22, 2019

 33 



[See all from Grant Holtes](#)[See all from DataSeries](#)

Recommended from Medium

 Cassie Kozyrkov

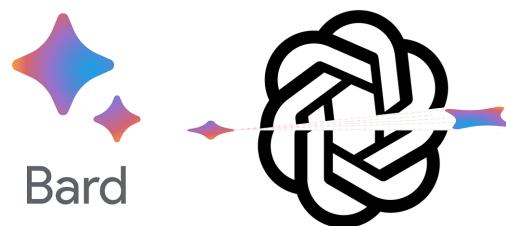
Why I quit my job as Google's Chief Decision Scientist

What's it like to go from being Chief Decision Scientist at Google to being, well... just me?

⭐ · 7 min read · Oct 28

 6.3K 90

...

 AL Anany 

The ChatGPT Hype Is Over—Now Watch How Google Will Kill...

It never happens instantly. The business game is longer than you know.

⭐ · 6 min read · Sep 1

 17.5K 532

...

Lists



Predictive Modeling w/ Python

20 stories · 560 saves



Practical Guides to Machine Learning

10 stories · 636 saves



Natural Language Processing

799 stories · 365 saves



The New Chatbots: ChatGPT, Bard, and Beyond

12 stories · 179 saves

{ JSON } is slow? ↘

Alternatives?

A screenshot of a terminal window shows the output of a command related to JSON performance.

Vaishnav Manoj in DataX Journal

JSON is incredibly slow: Here's What's Faster!

Unlocking the Need for Speed: Optimizing JSON Performance for Lightning-Fast Apps...

16 min read · Sep 28

👏 6.8K

💬 87



...

STATEMENT BY PRESIDENT
BARACK OBAMA
ON ISRAEL AND GAZA

October 23, 2023

Barack Obama

Thoughts on Israel and Gaza

It's been 17 days since Hamas launched its horrific attack against Israel, killing over 1,40...

5 min read · Oct 23

👏 37K

💬 921



...



Nima Tor... in Fundamentals of Product Manage...

The Inner Workings of Spotify's AI-Powered Music Recommendation...

Maximizing User Engagement: The Power of ML-Powered Recommendations at Spotify



Unbecoming

10 Seconds That Ended My 20 Year Marriage

It's August in Northern Virginia, hot and humid. I still haven't showered from my...

25 min read · Aug 28

★ · 4 min read · Feb 16, 2022

👏 12



+

...

👏 68K



988

+

...

[See more recommendations](#)