Analysis of Convolutional Neural Networks in Spotify's Music Recommendations

by Ronghe Chen

It's a Monday morning. You open up your newly-refreshed "Discover Weekly" playlist on Spotify, coming face-to-face with tunes that you have never heard before. Intrigued, you take a listen, and end up loving the songs that you have just discovered. You feel as if the Spotify algorithm has read your mind. This experience is made possible by the aggregation and analysis of data, powered by machine learning algorithms. Data analysis is crucial in the entertainment industry. By analysing user preferences and engagement metrics, companies gain insights into audience behavior, which helps them tailor their content towards their target audience. Spotify, established in 2006, is the biggest audio streaming service that holds about 35% of the market share (Maheshwari, 2023). Its success can be attributed to its recommendation system, which utilizes a variety of machine learning methods to collect and sort data pertaining to its users, their listening history, and the songs they've interacted with. The three methods used in Spotify's algorithm include collaborative filtering, natural language processing, and convolutional neural networks (Tambekar, 2020). Collaborative filtering uses matrix math to make automated predictions about the preferences of users based on their listening history and the preferences of similar users. Spotify's natural language processing algorithm searches the web for information about specific songs, using keywords to help them come up with a specific profile for those songs. And finally, songs are converted into raw audio files and their attributes are analysed by Spotify's convolutional neural networks, later being matched with songs with similar attributes. Convolutional neural networks (CNNs) are a type of deep learning model that use data for visual classification, where in this case, the waveforms of raw audio files are the input. The role that convolutional neural networks play within Spotify's recommendation algorithm is especially significant because it makes sure that less popular songs are considered by the algorithm, thus increasing accuracy. Due to the "big data" nature of collaborative filtering and natural language processing, more obscure songs may be obscured by these models. However, the nature of Spotify's convolutional neural networks focuses more on the physical attributes of a song rather than how it interacts with users. Spotify's use of convolutional neural networks in its algorithm has not only allowed users to discover new songs based off of their preferences, but also helping artists find their platform and strengthening Spotify's brand identity.

Convolutional neural networks are a form of deep learning with multiple layers. They are similar to artificial neural networks (ANNs), which are computational systems that model the human brain (O'Shea & Nash, 2015). However, a major difference between CNNs and ANNs is that CNNs are most often used for image recognition. Some specific applications of CNNs include facial recognition, automated driving, and medical imaging (*What Is a Convolutional Neural Network?*). The architecture of a CNN consists of many layers, an input layer, an output layer, with many hidden layers in between. The input layer takes in an image, which is later passed into the convolutional layer, which contains many kernels. The kernels "fire" when they see a specific feature at a given position in the input, and this process is called "activation". A rectified linear unit (ReLu) performs the activation function on the output of the previous layer, allowing for faster and more effective training. The pooling layer simplifies the output by reducing the dimensionality of the representation, as well as the number of parameters needed for the network to learn. There are often tens or hundereds of these layers between the input and output laters, each of which detect a certain feature in an image. As they get deeper and deeper into the model, the complexity of the features that they detect increases (see Figure 1).

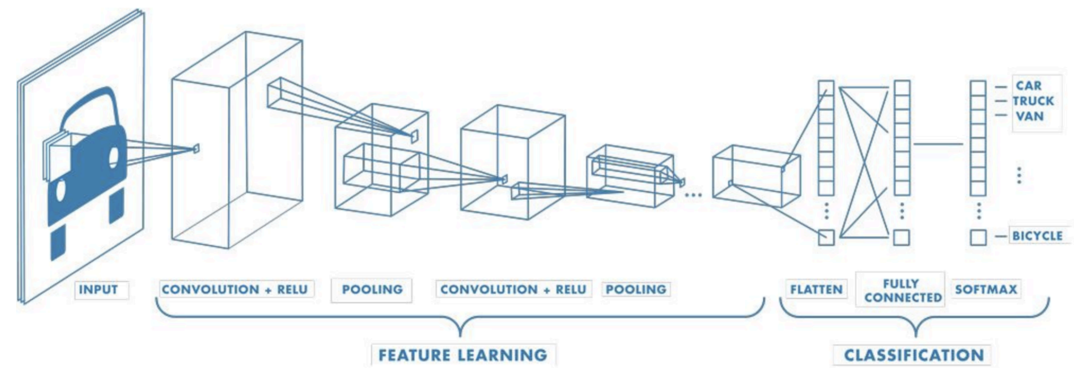

Figure 1

Layers of a CNN

However, Spotify uses them a little differently, applying them to sound instead of vision. Each song is converted to a raw audio file before being passed into the CNN, where they are represented as mel-spectograms (time-frequency representations). Bands slanting up and down in spectograms represent rising and falling pitches, and the CNN model's filters can detect human voices. The convolutional layers do the filtering, however, in this scenario, the two axes "time" and "frequency" do not have the same meaning, which is not the case for images. Between these layers, max-pooling operations are used to downsample the immediate representations in time and to add some invariance to the process. According to Sander Dieleman (2014), a former intern at Spotify who is now a researcher at Google DeepMind, the use of rectified linear units (ReLus) leads to faster convergence and reduces the vanishing gradient problem common in traditional neural networks. Parallelization is used to speed up training so that larger models can be trained in a shorter period of time. After the last convolutional layer, a global temporal pooling layer is added. This layer pools across the entire time axis, calculating statistics of the learned attributes across time. Three different pooling functions are used - mean, max, and the

L2 norm (see Figure 2). The songs passed in as inputs are later analyzed for key parameters, such as the pitch, tempo, or mood. These parameters are later being translated into numerical values, so that they can be measured. For example, the valence of a track is represented as a number from 0.0 to 1.0, where a higher value means a "happier" undertone to the track (*Spotify Web API Reference*). Similarly, energy is measured on a scale from 0.0 to 1.0, where a higher value means that the track is more energetic (i.e. fast, loud, and noisy). The tempo of a track is simply measured in beats per minute (bpm), and the key is represented by an integer used in standard pitch class notation.

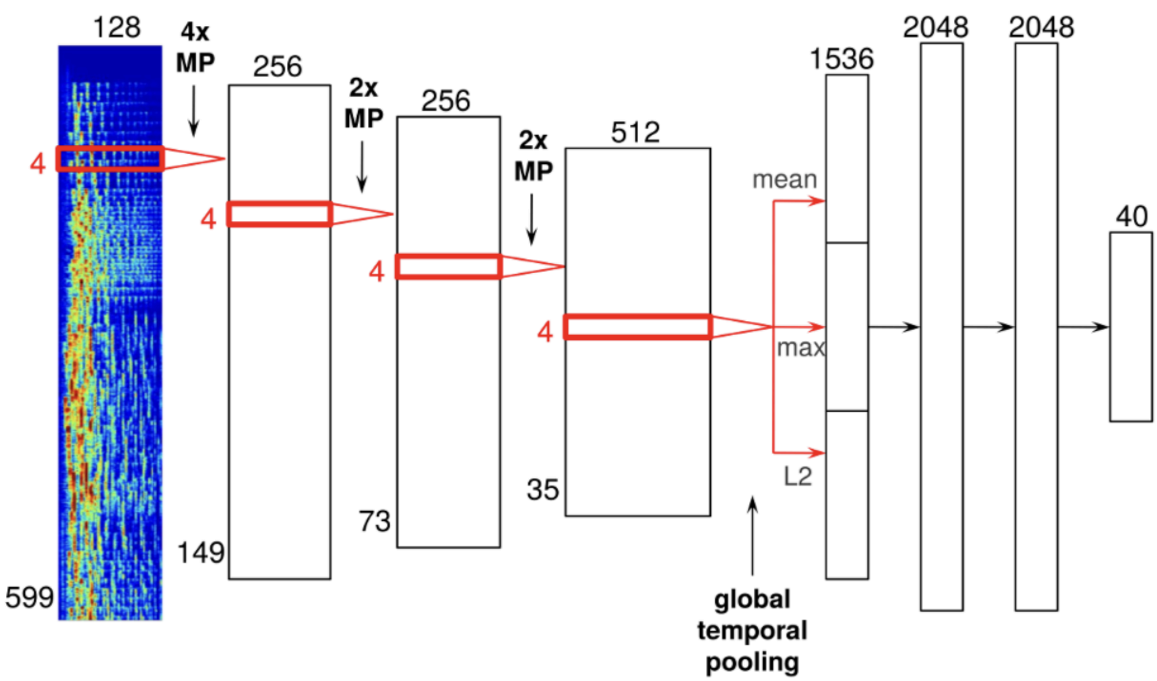

Figure 2

Spotify CNN tested by Sander Dieleman

Between July 2015 and July 2020, Spotify listeners have streamed over 2.3 billion hours of their personalized Discover Weekly playlist, so it is safe to say that Spotify's algorithm has received a fair amount of engagement over the years (Spotify Advertising, 2019). The engagement that they have received has been fueled by social media, as Spotify's advertising team reports seeing tweets where listeners say that their Discover Weekly playlists know them better than their spouses know them, or even in some cases, their playlists know them better than they know themselves. Another factor that has contributed to the success of Spotify's algorithm is brand loyalty. Discover Weekly users stream twice as long as non-Discover Weekly users, which has built up a lot of trust between Spotify and its users, thus allowing Spotify's audience to become more receptive to its message. Due to the detail-oriented nature of CNNs, rather than relying on big data alone. Spotify's algorithm has also helped lesser-known artists find their audience, with 16 billion artist discoveries every month on Spotify. However, the algorithm isn't perfect. Quartz editor Adam Pasick (2015) has said that his playlists usually contain one or two songs that he really loves, about half a dozen that he likes a lot, and a few "stinkers". Figure 3 shows a diagram of his music tastes, where his top genres include funk/soul and indie folk/rock, with one noticeable outlier - kids songs (for his daughter). Pasick asked Matthew Ogle, founder of now-deactivated startup This is My Jam, and Edward Newett, lead engineer for Spotify's Discover Weekly, how users can fine-tune their Discover Weekly results. They recommended users to skip songs they don't like, since users fasting forward within the first 30 seconds of a song is seen as a "thumbs down" by the algorithm, to explore new genres and music, and to add songs they like to a playlist, which gives Spotify more information about the preferences of the user. Other ways for users to influence their Discover Weekly results include using private mode, since Spotify's algorithm ignore what users listen to on private mode, and getting music recommendations from their friends.

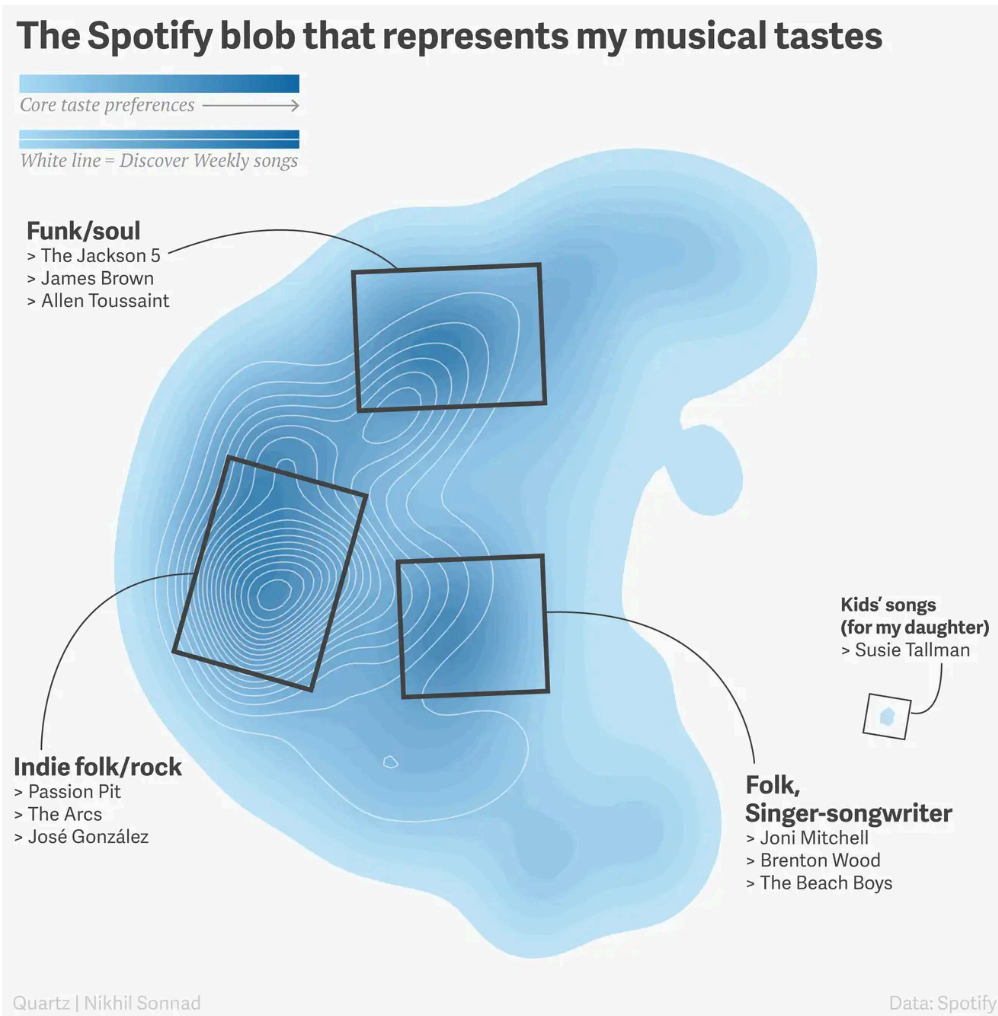

Figure 3

Diagram showing Adam Pasick's music taste, according to Spotify

Despite its massive outreach, Spotify's algorithm has not been met with unanimous approval. Critics believe that trying to quantify attributes of music is an oversimplification of music as art (Chodos, 2019). For example, "valence" is a measure of how positive or negative the emotionality of a song is. People who have devoted their lives to art and music believe that reducing the emotionality of a song to a number or a simple "good/bad" dichotomy does an injustice to the nuance and artistic nature of music. In addition to that, they argue that the perception of valence as either positive or negative could potentially lead to implicit bias in the system, and that human logic and behavior can never fully be represented by symbolic logic and brute computational force. Will machines and computers ever truly understand humans, or gain consciousness like that of humans that will allow them to experience music as an art? Researchers believe that the answer is no (Dehaene et al, 2017), since the computations used in deep learning (including CNNs) mostly model non-conscious operations in the brain. Consciousness is defined as global availability, the relationship between a cognitive system and an object of thought (i.e. thinking about something specific and being able to recall it, act upon it, or speak about it), and self monitoring, the ability for an entity to process and obtain information about itself (i.e. introspection or meta-cognition). Examples of self-monitoring in humans include knowing the position of their body, knowing whether they perceive something, or knowing if they had made a mistake. The computations performed by CNNs do not meet either of these criteria, since they merely recognize images and patterns. In humans, recognition of such stimuli without being able to recall or relate to it is not enough to activate brain processes relating to consciousness. Therefore, machines and algorithms do not have the capacity to experience music as an art, or to add conscious meaning to musical attributes, that humans have. Furthermore, different users have different preferences. Some listeners prefer songs with more danceability and energy, while others prefer more mellowed-out songs. In fact, some genres are more popular in different countries as a whole - for example, pop being more popular in China, Japan, and Indonesia, while metal is more popular in Finland, Bulgaria, and Turkey (Schedl et al., 2017). In addition to that, there is a relationship between age, personality, and music taste - for example, young adults between 20 and 39 prefer electronic music, while punk and alternative music are popular among adolescents high in neuroticism (Ferwerda et al., 2017). It is entirely possible that Spotify's recommendation system may be more effective for some demographics than others. Another example that demonstrates the nuance behind music recommendations is a 2018 study that had participants create a playlist of 9 songs under different situations (Millecamp et al., 2018). First, the participants select their favorite artists, then they manipulate musical attributes and select 9 songs from the recommendations. Finally, they answer a set of evaluation questions. It was found that having the ability to control musical attributes have allowed the participants to discover songs that have previously been ignored by either themselves or the system, which comes to show that the system is not perfect.

Spotify's use of convolutional neural networks on songs, where their attributes are represented as numerical values and later processed into the algorithm to be matched with songs with similar attributes, has contributed greatly to its success as a brand. It has not only helped users discover songs that they enjoy, including less popular ones that may have been neglected by other machine learning models, but it has also helped up-and-coming musicians gain more recognition. However, Spotify's algorithm has also faced a decent amount of criticism, with some people claiming that the value of music as an art cannot be reduced to numbers and computations, since machines and algorithms do not have consciousness the same way that humans do. Other criticisms pertain to different users having different preferences, and that the recommendation system would be more effective if users had more control over their preferences. The goal for machine learning in music recommendations is to promote a more holistic experience, focusing on recommendations that provide long-term satisfaction and enjoyment. A way to improve music recommendation algorithms is to design an interface that allows for users to have more control over their music recommendations, fine-tuning what specific attributes in music that they look for. In addition to that, recommendation systems must be able to take in an input and display a visualisation of the relationship between the input song and the recommended song, showing what musical attributes they have in common. According to Oskar Stål, Spotify's Vice President of Personalization, Spotify's objective is to use machine learning models to promote long-term gratification (Spotify, 2021). To do so, the model tries to predict how satisfied the user is while also nudging them to content that aligns with their preferences. Machine learning algorithms can be programmed to monitor learning progress - which is a technique called reinforcement learning. Although machines are still largely unconscious, the primary calculations can be performed with classical CNNs, while a second, hierarchically higher neural network is trained to predict the other network's performance. This kind of monitoring coupled with global consciousness could be a first step towards machines gaining a sense of consciousmess, where it has global access to parsed information and the sense that their content is a reflection of the immediate outside world. With machine learning used in tandem with reinforcement learning, Spotify aims to create a more holistic audio experience, maximising long-term enjoyment. In a world where technology is known to be cold and logical (Torabi, 2024), Spotify's algorithm revolutionizes the way humans interact with the precision and art of music, staying true to its brand value of a personalised experience.

References

Biancofiore, G. M., Di Noia, T., Di Sciascio, E., Narducci, F., & Pastore, P. (2022). Aspect based sentiment analysis in music. *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*. https://doi.org/10.1145/3477314.3507092

Chodos, Asher Tobin. (2019). Solving and Dissolving Musical Affection: A Critical Study of Spotify and Automated Music Recommendation in the 21st Century. https://www.proquest.com/openview/710170eea7ab205bcc705d534a549873/1?cbl=18750&diss=y&parentSessionId=xkJ%2BvhoHJL1XLivpaCAs31Da%2Bj6Yu9jTHGoGZCrFRj8%3D&pq-origsite=gscholar

Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it? *Science*, *358*(6362), 486–492. https://doi.org/10.1126/science.aan8871

Dieleman, Sander (2014, August 5). *spotify_convnet* [Photograph]. spotify_convnet.png

Dieleman, Sander. (2014, August 5). *Recommending music on Spotify with deep learning. https://sander.ai/2014/08/05/spotify-cnns.html#contentbased*

Ferwerda, B., Tkalcic, M., & Schedl, M. (2017). Personality Traits and Music Genre Preferences : How Music Taste Varies Over Age Groups. Proceedings of the 1st Workshop on Temporal Reasoning in Recommender Systems (RecTemp) at the 11th ACM Conference on Recommender Systems, 1922, 16–20. Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:hj:diva-36884

García Vázquez, F. (2017). *"Recommendation of songs through deep learning techniques" Esta obra se encuentra sujeta a la licencia Creative Commons Reconocimiento -No Comercial -Sin Obra Derivada*. https://core.ac.uk/download/pdf/288501631.pdf

O'Shea, K., & Nash, R. (2015). *An Introduction to Convolutional Neural Networks*. https://arxiv.org/pdf/1511.08458

Maheshwari, C. (2023). *Music Recommendation on Spotify using Deep Learning*. https://arxiv.org/pdf/2312.10079

Millecamp, M., Htun, N. N., Jin, Y., & Verbert, K. (2018). Controlling Spotify Recommendations. *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*. https://doi.org/10.1145/3209219.3209223

Pasick, A. (2015, December 21). *The magic that makes Spotify's Discover Weekly playlists so damn good*. Quartz. https://qz.com/571007/the-magic-that-makes-spotifys-discover-weekly-playlists-so-damn-good

Pasick, A. (2015, December 21). *Untitled* [Photograph]. Quartz. 97fb54c1adce1821fcb379d15d0718a2.jpg

Schedl, M., Florian Lemmerich, Ferwerda, B., Skowron, M., & Knees, P. (2017). Indicators of Country Similarity in Terms of Music Taste, Cultural, and Socio-economic Factors. *KTH Publication Database DiVA (KTH Royal Institute of Technology)*. https://doi.org/10.1109/ism.2017.55

Spotify. (2021, December 2). *How Spotify Uses ML to Create the Future of Personalization*. Spotify Engineering. https://engineering.atspotify.com/2021/12/how-spotify-uses-ml-to-create-the-future-of-personalization/

Spotify. *Web API Reference*. developer.spotify.com. https://developer.spotify.com/documentation/web-api/reference/get-audio-features

Spotify Advertising. *Five Years Of Discover Weekly*. (2019). ads.spotify.com. https://ads.spotify.com/en-GB/news-and-insights/five-years-of-discovery-and-engagement-through-discover-weekly/

Tambekar, A. (2020, May 1). *How Spotify Uses Machine Learning Models to Recommend You The Music You Like*. GreatLearning Blog. https://www.mygreatlearning.com/blog/how-spotify-uses-machine-learning-models/

Torabi, N. (2024, April 5). *The Inner Workings of Spotify's AI-Powered Music Recommendations: How Spotify Shapes Your Playlist*. Medium. https://neemz.medium.com/the-inner-workings-of-spotifys-ai-powered-music-recommendations-how-spotify-shapes-your-playlist-a10a9148ee8d#7e2e

Untitled [Photograph]. uk.mathworks.com. 1710246105363.jpg

Van Den Oord, A., Dieleman, S., & Schrauwen, B. (2013). *Deep content-based music recommendation*. https://papers.nips.cc/paper_files/paper/2013/file/b3ba8f1bee1238a2f37603d90b58898d-Paper.pdf

What are Convolutional Neural Networks (CNNs)? (2021). www.youtube.com. https://www.youtube.com/watch?v=QzY57FaENXg

What Is a Convolutional Neural Network? uk.mathworks.com. https://uk.mathworks.com/discovery/convolutional-neural-network.html