Modelo de Predicción de Éxito de Canciones Basado en Descriptores de Audio

Ignacio Barrera and Rodrigo Nazar
Pontificia Universidad Católica de Chile
Second Semester 2021
Link a Carpeta Drive: https://drive.google.com/drive/folders/17DjY0REw1XG8hWYBg9LRwIihbpYro6T9

***Abstract*—El presente trabajo busca modelar la predicción del nivel de éxito de una canción según el ranking de canciones más escuchadas en la plataforma Spotify. Se utiliza una metodología de SOCP para minimizar la región factible para los descriptores de audio de canciones, con lo que se generan restricciones para que una nueva canción sea similar a las utilizadas como datos. Se muestran una simulación computacional basada en datos de Spotify y de elaboración propia. Se concluye presentando el trabajo futuro que resta para la investigación.**

I. Introducción

A. Motivación del Problema

En la búsqueda de la construcción de la identidad propia, el ser humano ha procurado separar aquellas cosas que son de su agrado de las que no lo son, ya sea que le son indiferentes o le resulten molestas. En este ámbito, la música es un gran ejemplo de estas clasificaciones inconscientes que realizamos a diario. Los factores que determinan las preferencias son de lo más variados, y discutiblemente pueden llegar a ser considerados imposibles de modelar en su totalidad. Sin embargo, a nivel colectivo es posible notar la preferencia sobre ciertas composiciones, las cuales, aún sin poder describirse en su totalidad, son abordables de forma matemática con más facilidad que al hacerlo de forma individual. En particular, se pueden observar ciertas tendencias en cuanto a géneros musicales, ritmos y otros factores, así como también de otras no apreciables sin herramientas matemáticas y computacionales. El presente trabajo busca encontrar la mejor forma de seleccionar y combinar esas herramientas, en conjunto a los parámetros que las controlan, para modelar el nivel de popularidad que podría tener una canción.

B. Revisión Bibliográfica

Se entiende como descriptores o características de audio al resultado de la aplicación de algoritmos computacionales, que toman como entrada una señal de audio y entregan uno o más valores numéricos. El propósito de un descriptor es que este modele de manera correcta un fenómeno que se desea estudiar, por eso son un pilar fundamental en el análisis de señales.

Existen muchas formas de clasificar los descriptores. Un ejemplo es el dominio en el cual se operan los datos: procesar la señal en el tiempo, su espectro (la magnitud de su transformada de *Fourier*) u otra proyección a otro dominio. Otro ejemplo de clasificación es si son completamente autónomos o si necesitan parámetros externos que regulen su comportamiento. En el libro de Weihs [1] se presenta la teoría del análisis de señales basado en descriptores.

En cuanto a la descripción de tendencias la literatura más cercana se encuentra en los sistemas recomendadores, puesto que en esencia, una recomendación acertada de cualquier clase consiste en modelar correctamente las preferencias de un individuo respecto a elementos de algún conjunto y predecir si un nuevo item será de su agrado. Al respecto, las aproximaciones al problema son variadas. Maldonado y López [5] utilizan un SOCP para crear un clasificador para una Support Vector Machine. Otro algoritmo usado se encuentra en la descomposición por valores singulares, como la realizada por Mehta [6], en la cual aprovecha la ventaja de este método para reducir la cantidad de datos usados y para seleccionar solo aquellos que son más relevantes, con lo cual genera una recomendación.

Por otro lado, al momento de describir qué es lo que vuelve a una canción popular, Askin y Mauskapf [7] describieron 8 categorías en base a las que describen a una canción. Según los autores, estas categorías permiten encontrar un perfil de popularidad que se mantiene relativamente consistente dependiendo del año. Estas categorías son su danzabilidad, tempo, energía, acusticidad, vivacidad, discurso, instrumentalidad y valencia. Es importante notar que todos estos corresponden a parámetros ”superficiales”, en el sentido de que abordan la forma en que se aprecia la canción, y no su estructura propiamente tal.

C. Preguntas de Investigación

En base a lo anteriormente expuesto, la pregunta que se buscará responder es: ¿Qué requisitos debe cumplir una canción para estar incluida en la lista de ”éxitos”?

Además, una vez obtenida una modelación de este, se intentará conseguir un método para aproximar la posición en el ranking de una canción en función de las restricciones obtenidas al optimizar y sus descriptores

II. Descripción del Problema

Se utilizará la base de datos de Spotify para obtener el top 100 de las canciones más escuchadas a nivel mundial en el año 2017. Además, la cantidad de descriptores será evaluada durante el desarrollo del proyecto, más como referencia son 13 los que se utilizan en la sección de experimentos con datos de Spotify. El objetivo será obtener la proporción de aquellas

características que permitan generar un poliedro que encierre las características mínimas (apreciables o no) para determinar el "éxito" de una canción. Cada canción será categorizada según su lugar en el ranking, siendo la cercanía al primer puesto el nivel de "éxito" de la canción. Como se puede intuir, cuanto menor sea el valor, más exitosa será, por lo que se buscará minimizar cierta métrica. Para abordar el problema se escoge usar Programación Cónica de Segundo Orden (SOCP), debido a la forma "descendente" que tendrán los descriptores respecto a su valor en el ranking (cuanto menor sea su lugar, su aporte a las restricciones será mayor). Otro motivo para usarla es debido a que permite usar restricciones cuadráticas, las cuales calzan en bastantes casos con las características. Una vez obtenidos los descriptores, se puede probar el algoritmo añadiendo los datos de alguna canción que no esté en la lista, y comprobar qué posición le corresponde según el algoritmo con la posición que entregan los datos en la realidad.

III. Modelo matemático

Se propone la implementación de un modelo SOCP para generar restricciones que determinen la elegibilidad de una canción. El modelo presentado es el siguiente

A. Notación

1) Indices
   - i = descriptor
   - j = canción
2) Conjuntos
   - $\mathcal{I}$: Conjunto de descriptores de audio extraidas de una canción
   - $\mathcal{J}$: Conjunto de canciones usadas como dato
3) Parámetros
   - $\alpha_{i,j}$ = Descriptor i-ésimo de la canción j-ésima
4) Variables
   - $x_{i,j}$ = Coeficiente del descriptor i-ésimo de la j-ésima canción
   - $p_{i,j}$ = Inputs del descriptor compuesto i-ésimo para la canción j-ésima
   - $f(p_{i,j})$ = Coeficiente del descriptor compuesto i-ésimo para la canción j-ésima

B. Modelo

$$\min \quad -\left(\sum_{i\in\mathcal{I}} x_i)\right) \tag{1a}$$

$$\text{subject to} \quad z_j \geq \sqrt{\sum_{i\in\mathcal{I}} (\alpha_{i,j} \cdot x_i)^2} \quad \forall j \in \mathcal{J} \tag{1b}$$

$$f(p_{r,s}) \geq x_r, r \in \mathcal{I}, \forall s \in \mathcal{J} \tag{1c}$$

$$x_i \geq 0. \tag{1d}$$

Es importante recalcar que el valor de la función $f(p_{i,j})$ será numérico, por lo que la restricción 1c será lineal, lo cual mantiene la forma de cono de segundo orden del problema. Adicionalmente, se usaron los índices r y s para referir a valores dentro del conjunto $\mathcal{I}$ y $\mathcal{J}$, pues se trata de solo alguno de ellos, y no la totalidad de los valores de i y j (específicamente a los que están asociados a los descriptores compuestos).

La Función objetivo corresponderá a la suma de todos los coeficientes asociados a un descriptor. Se busca maximizar dicho valor para hacer un cono más restrictivo, lo que equivale a minimizar su negativo. La restricción 1b es la restricción de segundo orden que determina los parámetros del cono que envolverá los datos, mientras que la restricción 1d asegura la positividad de los valores. Por último, la restricción 1c busca modelar aquellos descriptores que requieran en sí mismos la selección de parámetros, tales como los descritos en la revisión bibliográfica, los cuales serán ajustados para obtener el mejor descriptor para el problema. La selección de esta función será evaluada durante el desarrollo del proyecto, sin embargo, por el momento puede considerarse como una constante. Se considera además la posibilidad de añadir nuevas restricciones propias de los descriptores, las cuales deberán ser evaluadas una vez que se tenga el dataset definitivo.

IV. Experimentos computacionales

En esta sección se expone el método de resolución problema propuesto anteriormente. También se presentan los datos utilizados y los experimentos realizados.

A. Metología

1) Datos Utilizados: Los descriptores, *rankings* y parámetros que se utilizaron en los experimentos provienen de dos fuentes:

- Bases de datos externas:
  Se utilizó el conjunto de datos de *Top Spotify Tracks of 2017* [4]. Esta contiene una tabla con las 100 canciones más escuchadas de *Spotify* del año 2017, en conjunto a 13 descriptores de cada una, entre los que se encuentran algunos propuestos por [7]. Las características incluidas son: danzabilidad, energía, tono, volumen, modalidad (menor o menor, en el sentido armónico), presencia de palabras habladas, acusticalidad, instrumentalidad, vivacidad, valencia, tempo, duración y estructura de tiempo de los compases. Las interpretaciones de estos descriptores están descritas por McIntire [2].
  Con esto, se tiene una matriz de 100 filas que representan cada canción y 13 columnas de descriptores asociadas a estas. Cada fila ordenada con su posición en el *ranking* de canciones más escuchadas.
- Descriptores calculados:
  Se cuenta con una base de datos de 128 extractos de 30 segundos de distintas canciones. Se calcularon 24 descriptores a cada una. Donde 13 de estos descriptores vienen del cálculo de los coeficientes MFCC de la señal. Los restantes 11 descriptores son: el centroide, media, varianza, oblicuidad, kurtosis, valor RMS, amplitud máxima, razón de cruces por cero, centroide espectral, pendiente espectral y planicie espectral. Las interpretaciones de estos descriptores están descritas en [1].

Si se quiere ver más en detalle el significado de cada una, se puede consultar el siguiente libro:
El procesamiento de las señales de audio finalmente produce una matriz de 128 filas (cada una por canción) y 24 columnas de características.

2) Procedimiento: Los distintos pasos y ambientes que se utilizaron se muestran en el diagrama de bloques de la figura 1. Este muestra el flujo de pasos para el proceso de extracción de descriptores propios. El método que se utilizó con la base de datos de *Spotify* es análogo, sólo que no cuenta con los dos primeros pasos de lectura de la señal y la extracción de sus características, estas venían precalculadas. En la figura también se especifican los datos de entrada y de salida de cada paso.

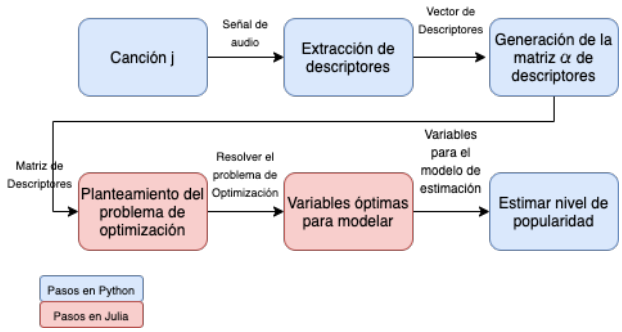

Fig. 1: Procedimiento del método.

3) Ambiente Computacional: La obtención de las características y manejo de los datos fueron programados en rutinas de *Python 3.7*. Por otra parte, el modelo de optimización y su solución fueron calculados con la librería *Ipopt* de *Julia 1.6*.

V. Resultados

Para ambos conjuntos de datos se obtuvieron los siguientes vectores de coeficientes óptimos:

$$x_{Spotify} = \begin{bmatrix} 0.091 \\ 0.141 \\ 0.057 \\ 0.006 \\ 1.523 \\ 9.277 \\ 0.188 \\ 104.3 \\ 6.895 \\ 0.072 \\ 6.485 \cdot 10^{-6} \\ 1.126 \cdot 10^{-12} \\ 0.0038 \end{bmatrix} \quad x_{Propios} = \begin{bmatrix} 0.0126 \\ 0.963 \\ 0.314 \\ 0.101 \\ 0.278 \\ 0.527 \\ 1.860 \\ 0.224 \\ 0.224 \\ 0.152 \\ 0.106 \\ 0.078 \\ 0.102 \\ 0.181 \\ 0.342 \\ 0.577 \\ 0.206 \\ 0.380 \\ 2.743 \\ 0.878 \\ 0.646 \\ 0.555 \end{bmatrix}$$

Cada vector representa los coeficientes que genera el cono más restrictivo que contiene los descriptores de las canciones. Con esto se busca delimitar la región en la cual se espera tener una canción exitosa, en base a sus descriptores.

A. Experimento

A continuación, se muestra el cono para una de las dimensiones de los descriptores. Específicamente, en la figura 2 se muestra el de la danzabilidad del top 100 de Spotify. Se puede apreciar que todos los valores de este descriptor son positivos, por lo que a futuro se necesitará adaptar el cono a dichos valores en lugar de hacerlo simétrico. Además, se observa que la pendiente es muy baja, esto debido a que la magnitud de dicho valor es baja. De esto se puede deducir que una metodología útil para el modelo puede ser normalizar cada valor. Una canción de prueba debería calzar dentro de los límites del cono, para poder considerarse "similar" y por lo tanto pertenecer a la tendencia del año en cuestión.

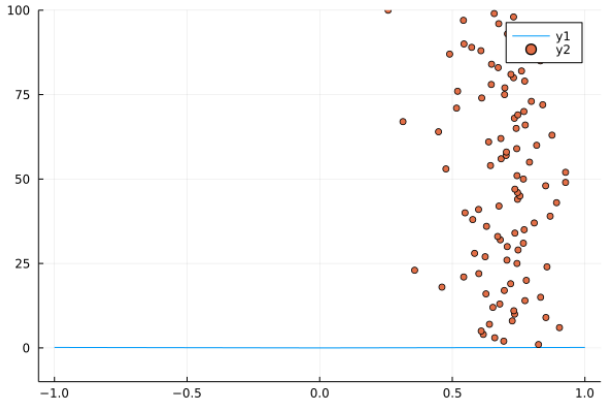

Fig. 2: Cono de la danzabilidad del top 100 de Spotify

VI. Trabajo Futuro

Como se señaló en la sección de experimentos, los datos utilizados no están completos, debido a que se está usando información previamente existente que no describe del todo bien el problema. Específicamente, se necesitará obtener los descriptores de elaboración propia para cada una de las 100 canciones. Estos descriptores ya se encuentran programados, por lo que solo resta conseguir el archivo de audio de las canciones y generar una nueva base de datos utilizando los códigos. Como señala [3], no será necesario extraer los descriptores de toda la canción, sino que basta con escoger 30 segundos representativos, debido a que por lo general, las canciones tienen muchos segmentos repetidos o similares en términos de su estructura.

Un segundo trabajo será escoger los descriptores más significativos. Esto será hecho mediante una descomposición de valores singulares, debido a que este método entrega, mediante el uso de valores propios, una intuición sobre la importancia de cada descriptor en la preferencia sobre una canción. En particular este método fue usado sobre los datos obtenidos de Spotify, donde se encontró que 6 de los 13 descriptores representan casi la totalidad de la información.

Como trabajo futuro queda la aplicación de este método sobre los descriptores de elaboración propia de las canciones, para luego incluir los más relevantes en el algoritmo de SOCP.

En cuanto a la función de la restricción 1d, esta deberá ser evaluada caso a caso, debido a la naturaleza de los descriptores. Algunos de los mencionados en la revisión bibliográfica requieren la selección de parámetros para funcionar, sin embargo su comportamiento no es explícito de ser modelado mediante una función, por lo que una alternativa será fijar valores, y al momento de realizar un análisis de sensibilidad, modificar aquellos que resulten ser más relevantes.

Otro tema abierto es la interpretación que debe dársele a las restricciones. Según [7], debe existir un trade-off entre la similaridad y diferenciación de una canción, por lo que si bien pertenecer al cono indicaría cierto nivel de éxito, estar cerca de la frontera (incluso si eso significa sobrepasarla) podría tener relación con el éxito de la canción. Estos factores deberán ser estudiados más en detalle para el avance del proyecto.

Finalmente, el trabajo de predecir la posición en el ranking quedará como trabajo futuro, debido a que se requiere la información obtenida de la optimización para realizar un modelo más acertado. Las alternativas que se barajan son una aproximación por mínimos cuadrados o un algoritmo genético, incluyendo como restricción que una canción, para ser considerada en el top 100, debe estar contenida en el cono de segundo orden.

VII. Distribución del trabajo

El trabajo fue colaborativo en casi todo el proceso. Rodrigo abarcó mayormente temas de programación, de obtención y tratamiento de datos. Mientras que Ignacio tocó temas de bibliografía y modelación, sin embargo, el trabajo fue realizado con ambos presentes, por lo que fue repartido e informado constantemente.

References

[1] C. Weihs, D. Jannach, I. Vatolkin, G. Rudolph. Music Data Analysis: Foundations and Applications. *Chapman & Hall/CRC*, ed. 2, pp. 147–162, 2017.

[2] G. McIntire. A Machine Learning Deep Dive into My Spotify Data, Version 1. Consultado el 20 de Septiembre, 2020 de *https://opendatascience.com/a-machine-learning-deep-dive-into-my-spotify-data/*, 2017.

[3] H. Huihui, X. Luo, T. Yang and S. Youqun. Music Recommendation Based on Feature Similarity. *IEEE International Conference of Safety Produce Informatization*, 2018.

[4] J. Song. Spotify DataSets, Version 1. Consultado el 20 de Septiembre, 2020 de *https://www.kaggle.com/jsongunsw/spotify-datasets*, 2017.

[5] S. Maldonado and J. Lopez. Imbalanced data classification using second-order cone programming support vector machines. *Pattern Recognition*, vol. 47, no. 5, pp. 2070–2079, 2014.

[6] R. Mehta. Session Based Music Recommendation using Singular Value Decomposition (SVD). *San Jose State University Scholar Works*, Master's Projects. 237., 2012.

[7] N. Askin and M. Mauskapf. What Makes Popular Culture Popular? Product Features and Optimal Differentiation in Music. *American Sociological Review*, vol. 82, no. 5, pp. 1-35, 2017.