División de Ciencias Básicas e Ingeniería
Posgrado en Ciencias y Tecnologías de la Información

Reconocimiento de Géneros Musicales Aplicando Técnicas de Aprendizaje Maquinal.

Idónea Comunicación de Resultados que presenta:
Lic. Miguel Ángel Ramírez Gómez

Para obtener el grado de
Maestro en Ciencias y Tecnologías de la Información

Dirigida por:
M. en C. Fabiola M. Martínez Licona

Ciudad de México
7 Julio 2016

RECONOCIMIENTO DE GÉNEROS MUSICALES APLICANDO TÉCNICAS DE APRENDIZAJE MAQUINAL.

Idónea Comunicación de Resultados que presenta:

Lic. Miguel Ángel Ramírez Gómez

Para obtener el grado de

Maestro en Ciencias y Tecnologías de la Información

Dirigida por:

M. en C. Fabiola M. Martínez Licona

Defendida públicamente en la UAM Iztapalapa

El 7 de Julio del 2016 a las 12:00 hrs.

JURADO:

Dr. José Francisco Martínez Trinidad

INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA

PRESIDENTE

M. en C. Alma E. Martínez Licona

UNIVERSIDAD AUTÓNOMA METROPOLITANA IZTAPALAPA

SECRETARIO

M. en C. Fabiola M. Martínez Licona

UNIVERSIDAD AUTÓNOMA METROPOLITANA IZTAPALAPA

VOCAL

Abstract

The intersection between music, machine learning and signal processing has allowed addressing a wide range of tasks such as automatic identification of songs, instruments, genre or artist. In particular, the identification of musical genres is a technique that is used in most digital playback systems for managing music songs that are archived on their mechanisms. Given the large number of tracks that can be stored, the location of a particular song can become complicated if the song set is not properly organized; the generation of playlists requires categorized systems with a high confidence degree of the songs to choose from.

A Musical Genres Recognition System (MGRS) requires a set of musical elements, a selection of features that describe the musical genres and a method of classification. The task is complicated by the limited availability of data (songs) for use in research because of intellectual property issues. In this project the performance of different machine-learning methods to define a MGRS, including those based on the theory of deep belief was compared.

The database Million Song Dataset (MSD) was used for experimentation and the musical genres to recognize were chosen based on previously established criteria. A pre-processing was performed to generate a vector of smaller dimension attributes with the most relevant features of the database. There were analyzed and experimented different techniques of mechanical learning as K-means, perceptron, support vector machines and deep belief networks.

It was found that timbre, pitches and intensity characteristics were those which offered good results; in particular the first 4 timbre coefficients and the application of the average feature vectors improved ranking results in several cases. The best performing system offered was composed of a deep belief network with mean and standard pitch, timbre and intensity as attributes on a set of 6 musical genres as classes.

The use of attributes taken from a descriptive database as it is MSD allowed to appreciate the contribution of acoustic features on musical genre recognition, which gives the idea of continuing with experimentation in direct audio data.

Resumen

La intersección entre la música, el aprendizaje maquinal y el procesamiento de señales ha permitido abordar un amplio rango de tareas como la identificación automática de canciones, instrumentos, género o artista. En particular, la identificación de géneros musicales es una técnica que utiliza la mayoría de los sistemas de reproducción de música digital para gestionar las canciones que los mecanismos guardan. Dado el gran número de piezas musicales que se puede almacenar, la localización de una canción en particular puede llegar a ser complicada si el conjunto no se encuentra organizado adecuadamente; la generación de listas de reproducción requieren de sistemas que categoricen con un alto grado de confianza las canciones a elegir.

Un Sistema de Reconocimiento de Géneros Musicales (SRGM) requiere de un conjunto de elementos musicales, una selección de características que describan los géneros musicales y un método de clasificación. La tarea se dificulta por la limitada disponibilidad de los datos (canciones) para su uso en investigación debido a aspectos de propiedad intelectual. En este proyecto se comparó el rendimiento de diferentes métodos de aprendizaje maquinal para definir un SRGM, incluyendo los basados en la teoría de la creencia profunda.

Se utilizó la base de datos Million Song Dataset (MSD) para la experimentación y se eligieron los géneros musicales a reconocer con base en criterios establecidos previamente. Se llevó a cabo un preprocesamiento para generar un vector de atributos de menor dimensión con las características más relevantes de la base de datos. Se analizaron

y experimentaron diferentes técnicas de aprendizaje maquinal como K-medias, perceptrón, máquinas de soporte vectorial y redes de creencia profunda.

Se encontró que las características timbre, pitches e intensidad fueron las que ofrecieron buenos resultados; en particular los primeros 4 coeficientes del timbre así como la aplicación de la media a los vectores de características mejoraron los resultados en la clasificación en varios casos. El sistema que mejores resultados ofreció fue el compuesto por una red de creencia profunda con los atributos de medias y varianzas normalizadas de pitch, timbre e intensidad sobre un conjunto de 6 géneros musicales como clases.

El uso de atributos tomados de una base de datos descriptiva como lo es MSD permitió apreciar el aporte de las características acústicas al reconocimiento de géneros, lo cual da la idea de seguir con la experimentación en datos directos de audio.

Agradecimientos

Agradezco al Posgrado en Ciencias y Tecnologías de la Información y a la UAM Iztapalapa por darme la oportunidad de desarrollarme de manera personal y profesional con la realización de este proyecto. Así mismo agradezco al CONACyT por el apoyo económico brindado.

Agradezco a la Mtra. Fabiola Martínez Licona todo su apoyo y confianza, más que una guía fue un soporte muy importante para mí, revisó e indicó un sin número de mejoras para aclarar y dar fluidez a la investigación.

Agradezco a la Mtra. Alma Martínez Licona de la UAM-I y al Doctor Francisco Martínez Trinidad del INAOE por haber aceptado fungir como revisores de esta Idónea Comunicación de Resultados, contribuyendo con valiosos comentarios para enriquecerla.

Agradezco a mis profesores y compañeros que compartieron su conocimiento en pro de nuestro desarrollo y finalmente a mi familia por su apoyo siempre incondicional.

Muchas Gracias.

Miguel Ángel Ramírez Gómez

Junio 2016

Índice general

Índice de figuras

Índice de tablas

The beginning is the most important part of the work.

Plato

CAPÍTULO

1

Introducción

La música es uno de los medios de comunicación más importantes ya que es capaz de despertar emociones y estados de ánimo en quien la escucha. El hombre primitivo percibió en los sonidos de la naturaleza y en su propia voz, los elementos que le permitieron utilizar objetos rudimentarios (huesos, cañas, troncos, conchas, etc.) para producir sonidos y comunicarse.

La música actualmente forma parte de nuestra vida diaria, por ejemplo, claramente es un elemento indispensable en la creación de contenidos para la televisión, radio e Internet. Además esto ha favorecido el incremento en número y tamaño de las bases de datos musicales. Continuamente se adicionan grandes cantidades de música en línea, debido principalmente a la publicación de música en Internet y a la restauración de archivos analógicos existentes gracias a los avances de las tecnologías web. En consecuencia, cada vez se requiere de herramientas más rápidas y fiables para el análisis de contenidos, recuperación y descripción de música, acceso interactivo, consultas de música basadas en contenidos, etc.

Los géneros musicales son etiquetas creadas por los seres humanos para caracterizar a la música y poderla clasificar. Un género musical se identifica por las

características comunes que comparten sus miembros (canciones); éstas normalmente están relacionados con la instrumentación, la estructura rítmica, y el contenido armónico de la música. El continuo surgimiento de géneros musicales y su combinación dificultan la tarea de clasificar las canciones por lo que habitualmente se utilizan jerarquías de género para estructurar las grandes colecciones de música disponibles en la Web.

Actualmente la notación del género musical se realiza manualmente de forma subjetiva, lo que conlleva problemas que van desde la asignación arbitraria del género hasta la confianza y precisión de la información que depende en gran medida de la experiencia y conocimiento del usuario.

El término recuperación de información hace referencia a tres actividades, la definición y organización de la información, la especificación de la búsqueda, y los sistemas y técnicas utilizadas para estos procesos. Jean Tague-Sutcliffe identifica los siguientes elementos que definen el campo de la recuperación de la información: a) colección de documentos o bases de datos, b) representación de la información, c) usuarios, d) consultas y estrategias de búsqueda (frases, oraciones,...), e) intermediarios de búsqueda, f) proceso de búsqueda, y g) evaluación de la recuperación [1]. No todos estos aspectos tienen importancia para los investigadores, algunos se centran en la entrada del sistema (representación y almacenamiento de la información), otros en la salida (búsqueda) y otros en el propio sistema (diseño del sistema de recuperación de información).

En comunidades como el ISMIR (International Society for Music Information Retrieval) y MIREX (Music Information Retrieval Evaluation eXchange) se presentan investigaciones enfocadas en solucionar problemas relacionados con la recuperación de información enfocada en la música, donde se contempla el reconocimiento de géneros musicales entre otras tareas.

Los géneros musicales son particularmente importantes porque expresan la identidad general de las culturas en las que se incorporan [2]. Además, forman parte de la interacción entre culturas, artistas y estrategias de mercado definiendo asociaciones

entre el artista y sus obras. El reconocimiento de géneros musicales puede ayudar a complementar lo que se conoce acerca de las trayectorias de los géneros de la música, su historia y su dinámica [3].

Es de suma importancia el impacto que puede tener la música en la sociedad y en las personas. La investigación en Recuperación de Información de la Música (MIR por sus siglas en inglés) puede ser de gran ayuda, por ejemplo, al proveer al usuario de tecnologías capaces de generar o recomendar música automáticamente, de esta forma el usuario podría escuchar música acorde para mejorar su estado de ánimo o de salud. La predicción de tendencias podría ayudar a la industria a identificar si un tema será un hit o si no va a producir ganancias. El mercado musical se beneficiaría de la información geográfica, étnica y social que se desprende de los sistemas que socializan los gustos musicales y así podríamos encontrar más casos.

Un Sistema de Reconocimiento de Géneros Musicales (SRGM) requiere principalmente de un conjunto de elementos musicales (canciones) sobre el cual trabajar, una selección de características que describan los géneros musicales y por supuesto, un método de clasificación. Si se logra cubrir estos requerimientos, se contará con un sistema que, al seleccionar una canción y a través de los procedimientos de aprendizaje maquinal que existen o de una combinación de ellos, será capaz de determinar correctamente el género musical al que corresponda.

El reconocimiento del género musical realizado por un humano es una tarea fácil para ciertos géneros, sin embargo, para una computadora no es sencillo. Para abordar este problema se han desarrollado investigaciones, donde los esfuerzos se han centrado en diferentes métodos utilizados para la extracción, análisis y selección de características de las canciones. Éstos se aplican antes de la etapa de clasificación la cual ya está muy desarrollada con métodos muy consolidados como los árboles de decisión, K-vecinos cercanos y las máquinas de soporte vectorial [4].

En cuanto al conjunto de elementos para el SRGM, una limitante es la disponibilidad y el tamaño de las bases de datos (BD). Las BD's que suelen estar disponibles son pequeñas mientras que las grandes BD's o no existen o tienen disponibilidad

limitada. Esto se debe a la falta de concesiones de las licencias necesarias para el uso adecuado de las canciones para fines de investigación.

La BD que se utilizará en este proyecto cuenta con un millón de canciones. Esto resuelve el problema de contar con datos suficientes, pero ahora nos enfrentamos a un problema de big data. Al tener una gran cantidad de datos, existen problemas inherentes como el almacenamiento de la información, que se puede resolver con las tecnologías actuales. Sin embargo el problema principal se puede resumir en cómo agilizar el acceso a estos datos de forma que sean mayormente redituables e imprescindibles para que el procesamiento, manejo y análisis de estos sea más valioso para nuestro problema.

Cuando las BD's llegan a ser muy grandes, los algoritmos comúnmente utilizados, requirieren demasiado tiempo para procesar la información, por lo que el problema se vuelve intratable. Una forma de atacar este problema es reducir los datos antes de aplicar tales algoritmos [5]. La selección de atributos aplicada como una etapa de pre-procesamiento a la minería resulta útil pues busca eliminar los atributos irrelevantes y/o redundantes sin afectar la calidad de la clasificación que realiza el algoritmo minero. De esta manera el porcentaje de ejemplos correctamente clasificadas llega a ser más alto, pues los datos a tratar quedan libres de ruido o de datos que provocan la generación de más nodos que los necesarios [6].

Para este tipo de problema el uso de una red de creencia profunda (DBN por sus siglas en inglés) puede resultar de utilidad ya que la estrategia del algoritmo es lograr la fusión de las características de la señal de audio para que se generen características más relevantes para realizar la clasificación.

En este proyecto se comparó el rendimiento de diferentes métodos de aprendizaje maquinal para definir un SRGM incluyendo los basados en la teoría de creencia profunda. Se llevó a cabo un pre-procesamiento para generar un vector de características de menor dimensión con las características más relevantes de los elementos de la BD. Algunas características se eligieron a partir de las reportadas en la

literatura. Se analizaron y experimentaron distintas técnicas de aprendizaje maquinal (AM) para encontrar las arquitecturas que dieran los mejores resultados. Se realizaron pruebas sobre algunos sub-conjuntos de datos limitados en número de géneros musicales y ejemplos, a fin de llevar acabo los ajustes y evaluaciones para la optimización del rendimiento del SRGM. El rendimiento fue evaluado con base al porcentaje de ejemplos clasificados correctamente.

Personally, I think it does help, that it makes a beneficial difference, but the scientific literature on the subject is very messy.

Jeanne Petrek

CAPÍTULO

2

Antecedentes

Los investigadores se han involucrado en este campo porque aún no se resuelve la cuestión primordial de comprender cómo múltiples sonidos se relacionan entre sí formando señales de audio musicales complejas que transmiten contenidos mediante la elaboración de una estructura temporal [7] [8] [9] [10].

Aunado a esto, el agrupamiento por géneros musicales es un proceso subjetivo que resulta muy influido por el conocimiento personal, la forma de sentir y la forma de escuchar la música por parte del oyente. Esto se debe al alto grado de abstracción que se requiere para encontrar características comunes a un género.

Los géneros musicales también dependen de la ubicación geográfica donde surgen así como del contexto social y cultura donde se desenvuelven, es por eso que continuamente van surgiendo nuevos géneros. De esta manera el reconocimiento resulta una tarea compleja de realizar.

La intersección entre la música, el aprendizaje maquinal y el procesamiento de señales ha permitido abordar un amplio rango de tareas de recuperación de información de la música entre las que se encuentra el reconocimiento del género musical [11]. Sin embargo muchos de los trabajos considerados están limitados por la

BD que utilizan debido al problema de las licencias.

Muchas de las BD's de música disponibles cuentan con música de ciertas regiones, y por lo regular están limitadas en diversidad (géneros, zonas, años, artistas, etc.) lo que ha motivado la creación de la base de datos MSD (Million Song Dataset) [1] que consta de características y metadatos para un millón de canciones bajo la licencia Creative Commons (CC) [2] como apoyo a los investigadores [12], además de empujar los límites de la investigación de MIR a escala comercial.

MSD es una BD creada por LabRosa en colaboración con The Echo Nest disponible en línea con un peso total aproximado de 280 GB. MSD contiene características de audio y metadatos en formato HDF5 (Hierarchical Data Format) de canciones de música popular contemporánea. El formato HDF5 para archivos y librerías fue diseñado para almacenar y organizar las grandes cantidades de información y datos numéricos que contienen.

Existen diversas cuestiones que necesitan ser abordadas con cuidado, ya que representan retos importantes para la correcta clasificación de géneros musicales. Tal es el caso de la selección de características de las canciones, tanto descriptivas como acústicas, además de los métodos de extracción de características y los métodos de clasificación. El procesamiento de señales de música se enfoca en características acústicas específicas y estructurales tales como la melodía, la armonía, el ritmo y el timbre que las distinguen de la señal de voz u otras señales no musicales [10]. Las técnicas de procesamiento de señales juegan un papel importante, estas nos sirven para extraer una gran variedad de información y descripciones como los tonos[3], la polifonía[4], los timbres[5] y los pulsos[6], los cuales pueden ser importantes para

[1] Una mayor descripción de la base de datos, así como los enlaces para su descarga, se puede encontrar en http://labrosa.ee.columbia.edu/millionsong/

[2] http://creativecommons.org/about

[3] La preeminencia, privilegio o superioridad de periodicidades fundamentales distintas.

[4] La preponderancia, dominio o superioridad de la superposición de las fuentes de sonido en conjuntos musicales.

[5] La variedad de características de la fuente.

[6] La jerarquía regular de las estructuras temporales.

diferentes tipos de aplicaciones. Cuando se trabaja sobre un conjunto de géneros bien definidos y claramente distinguibles los clasificadores tradicionales funcionan bien. Sin embargo este es un caso especial muy poco frecuente.

En la literatura existen diversos estudios sobre reconocimiento automático de géneros musicales entre estas se encuentran investigaciones previas como la elaborada en [13] donde para la extracción de características utilizaron tres métodos de cadena[1] los cuales se comparan entre sí y con otros modelos n-grama[2] de características globales. Se destaca que las características globales relacionadas con el tiempo dan mejores modelos de clasificación que los basados en características relacionadas con el tono.

En [14] se propone un sistema automático de extracción de características de audio para la tarea de clasificación de géneros, compuesto por una DBN sobre la transformada discreta de Fourier del audio. En [15] se propone un algoritmo voraz para el aprendizaje rápido de una DBN. En [16] se propone utilizar las DBN para unificar las etapas de extracción de características y la de interpretación semántica, aprendiendo de forma automática características que pueden ofrecer una visión objetiva de la obra musical y de los atributos relevantes para una tarea determinada. Tal parece que la teoría de la creencia profunda presenta elementos atractivos para la tarea de clasificación de géneros musicales, aunque estos métodos tienen una fuerte dependencia del tipo de características con las que se van a clasificar las canciones.

En [17] se utiliza un esquema de ensamble dinámico a partir de clasificadores seleccionados de un conjunto que se generó para llevar a cabo la clasificación de géneros musicales utilizando K-vecinos más cercanos En [18] se utilizan características espectrales para identificar a través del ritmo diferentes géneros musicales considerando los fenómenos psico-acústicos de acuerdo a la percepción humana utilizando

[1]Los métodos de cadena se basan en una representación secuencial de la música que considera una pieza como una cadena de símbolos. Se calcula una medida de similitud por pares entre las cadenas y se utiliza para clasificar las piezas sin marcar.

[2]Un modelo de n-grama puede ser definido por una cadena de Markov de orden n-1. Por tanto los modelos de n-grama son modelos ocultos de Markov.

una máquina de soporte vectorial (SVM por sus siglas en inglés) para la clasificación.

En [19] se utilizan los espectrogramas (Transformada de Fourier de Tiempo Corto o STFT por sus siglas en inglés) de las canciones y se extraen características de esta representación visual utilizando algunos descriptores de GLCM (Gray Level Co-occurrence Matrix). Se utilizan métodos de zonificación, para obtener la información local de un patrón dado en cada zona con SVM como clasificador para después aplicar un esquema de mayoría de votos para obtener la clasificación final. En [20] a diferencia de [19], se utiliza el operador Patron Local Binario (LBP por sus siglas en inglés) así como 4 reglas para la decisión final: máximo, mínimo, producto y suma.

Los conjuntos de datos utilizados en cada uno de estos estudios son muy diferentes a MSD y entre sí:

La base de datos **Dance-9** utilizada en [13] es una gran colección de melodías populares limitada a Europa, se subdivide en nueve categorías o géneros dependiendo el tipo de danza, contiene 2198 melodías y el formato de los archivos es MIDI.

La base de datos **GTZAN** utilizada en [14][18] consta de 1000 clips de 30 segundos de audio para 10 géneros musicales. El conjunto de datos es equilibrado tiene 100 clips para cada género y se ha utilizado como una referencia para la tarea de reconocimiento de género (ver tabla 2.1). En [18] se utilizan las bases de datos de la tabla 2.1.

En [17][19][20] los experimentos se llevan a cabo en un subconjunto de **Latin Music Database** (LMD). La LMD se compone de 3227 piezas de música en formato MP3, distribuidas uniformemente a lo largo de 10 clases de géneros musicales: Axe, Bachata, Bolero, Forro, Gaúcha, Merengue, Pagode, Salsa, Sertaneja y Tango. En [17] se utiliza un subconjunto de 1,300 piezas musicales y en [19][20] se utiliza un subconjunto de 900 piezas. Además en [20] se hace uso de la base de

Tabla 2.1: Bases de Datos.

GTZAN	1000	ISMIRrhythm	698	ISMIRgenre	1458
blues	100	ChaChaCha	111	classical	640
classical	100	Jive	60	electronic	229
country	100	Quickstep	82	jazz blues	52
disco	100	Rumba	98	metal punk	90
hiphop	100	Samba	86	rock pop	203
jazz	100	SlowWaltz	110	world	244
metal	100	Tango	86		
pop	100	VienneseWaltz	65		
reggae	100				
rock	100				

datos **ISMIRgenre**.

Algunas de las tareas del MIR tales cómo reconocimiento de covers, detección del tono, identificación de artistas, por mencionar algunas, han utilizado MSD. La descripción de MSD se encuentra en el apéndice A.

Para este proyecto se buscó realizar una extracción de características y explorar las arquitecturas que presentan las técnicas de aprendizaje maquinal, incluyendo el aprendizaje profundo. [14][15][11][21].

Historical methodology, as I see it, is a product of common sense applied to circumstances.

Samuel E. Morison

CAPÍTULO

3

Metodología

En esta sección se presenta la metodología propuesta para el desarrollo del sistema de reconocimiento de géneros musicales del proyecto. En general consta de 3 etapas:

- obtención de datos y selección de características a utilizar,
- aplicación de técnicas de aprendizaje no supervisado y supervisado (redes neuronales incluyendo las de aprendizaje profundo) para la clasificación de géneros musicales
- realización de pruebas y ajustes para obtener el sistema que mejor rendimiento tenga. La figura 3.1 muestra la metodología empleada.

3.1 Obtener Base de Datos.

Se determinaron los criterios de selección tanto de los géneros musicales como de las canciones de MSD, apoyándose en los géneros utilizados en la literatura consultada. Cabe mencionar que la base de datos sobre la que se trabajó tiene una distribución desigual de sus canciones con respecto a los géneros así como que

Figura 3.1: Metodología

incluye géneros musicales cuya trascendencia es más local (por ejemplo música country o cristiana). De los géneros encontrados en la literatura se eligieron los géneros musicales que tuvieran perceptualmente hablando elementos que pudieran distinguirlos entre sí. El conjunto obtenido se dividió en tres subconjuntos: uno de entrenamiento, otro de prueba y finalmente uno de validación.

3.2 Análisis y Selección de Componentes.

Los atributos de la base de datos MSD incluyen valores numéricos y categóricos que abordan diferentes perspectivas de las canciones que contiene. Por ejemplo dentro de los atributos se pueden encontrar los valores de los pitches por segmentos elegidos para cada tipo de nota dentro de un vector denominado croma[1], las etique-

[1]Representa la inherente circularidad del tono en clases. Dos octavas de un tono están relacionadas si comparten el mismo ángulo (en el circulo del croma). Suma de la energía en las frecuencias

tas que se asocian con el artista de acuerdo a musicbranz[1] o un indicador numérico que determina qué tan bailable es la canción. En esta etapa se procedió a seleccionar los atributos que se utilizarían para el sistema de reconocimiento de géneros musicales; la estrategia consistió en eliminar los atributos cuya contribución impactara menos en el desempeño del clasificador en un conjunto de validación independiente. Un criterio adicional que se utilizó fue el seleccionar los atributos del conjunto cuantitativo que describiera acústicamente las canciones con la finalidad de minimizar los posibles sesgos que se pudieran encontrar en la base de datos.

En este sentido, una característica interesante a nivel de segmento es el timbre, el cual afecta a aspectos fundamentales de la música. Asimismo la extracción del ritmo, el tempo[2] y la línea de bajo podrían ser útiles por lo que se consideró el análisis del atributo *segments timbre* de manera particular. Este atributo de MSD representa información del timbre de cada pista como un número diferente de segmentos dependiendo de la longitud de la canción. Se representa por una matriz de 12 x N-segmentos, donde N varía muy ampliamente para cada canción (desde 400 a más de 1600) y representa las características textuales a partir de la codificación de una combinación de parámetros (MFCC+PCA-like). Este tipo de codificación permite evaluar diferentes estrategias de clasificación, mismas que se implementaron en el marco de referencia del aprendizaje maquinal.

3.3 Análisis y Selección de Técnicas de Aprendizaje Maquinal.

Esta etapa se procedió a desarrollarla en dos partes: primero se analizó la capacidad de agrupamiento de los datos elegidos a través de métodos de aprendizaje no supervisado, en particular del método de K-means. Debido a que la documentación

de cada clase de tono.

[1]Enciclopedia musical que colecciona metadatos y está disponible a todo público https://musicbrainz.org

[2]Tempo:Termino musical que hace referencia a la velocidad con la que debe ejecutarse una pieza musical.

relacionada con el género asociado a cada canción viene de una fuente distinta a MSD [22]. Esto implica que la clase de cada instancia que iría al clasificador dentro de la base de datos utilizada no viene incorporada por lo que, tomando como base los atributos acústicos elegidos en la etapa anterior, se analizó su capacidad de agrupamiento y se contrastó con las clases obtenidas de TuWien.
Asimismo se probaron otros métodos de aprendizaje supervisado como Perceptrón y Máquinas de Soporte Vectorial para observar el modo en que se agrupaban los datos de la base elegida.

Una vez que se analizaron los datos bajo el esquema supervisado y no supervisado, se procedió a definir la arquitectura del clasificador de géneros musicales con base en el aprendizaje profundo. Primero se utilizaron las máquinas restringidas de Boltzman en la etapa de adecuación de los datos; esto en sí corresponde a una clase de aprendizaje no supervisado que permite identificar la capacidad de agrupamiento de los datos; en el Anexo C se describe brevemente esta técnica. Cabe mencionar que esta parte se realizó en la etapa previa, pero con otros fines. Posteriormente se definió la arquitectura de la red de creencia profunda a partir de elecciones con bases heurísticas donde se incluye el número de capas y el número de nodos en cada capa, la tasa de aprendizaje, número de iteraciones y función objetivo.

Finalmente se compararon los resultados obtenidos de todas las técnicas aplicadas con la finalidad de obtener el SRGM y valorar los métodos de aprendizaje profundo para esta BD.

3.4 Pruebas, Ajustes y Evaluación.

Se realizaron los experimentos correspondientes a cada configuración y se ajustaron los parámetros con base en los resultados, en particular con la tasa de error (error rate) obtenida. Primeramente se utilizó el conjunto de entrenamiento, se analizaron los resultados obtenidos con el conjunto de atributos seleccionados y se hicieron los ajustes necesarios que permitieran mejorar la clasificación. Posteriormente se utilizó el conjunto de prueba para verificar el funcionamiento del sistema

y realizar los ajustes necesarios. Finalmente se utilizó el conjunto de validación para evaluar el sistema de reconocimiento de géneros musicales final.

3.5 Clasificador de Géneros.

Como resultado final se obtuvo un sistema de reconocimiento de géneros musicales con la configuración con la que se obtuvieron los mejores resultados.

The logic of validation allows us to move between the two limits of dogmatism and skepticism.

Paul Ricoeur

CAPÍTULO

4

Resultados

En esta sección se presentan los resultados más significativos para la investigación obtenidos de la experimentación de acuerdo a las etapas establecidas en la metodología.

4.1 Definir Géneros a Utilizar de MSD.

Los criterios utilizados para seleccionar los géneros de la base de datos fueron:

- su existencia en otras bases de datos utilizadas para el reconocimiento de géneros musicales (Tabla.2.1),
- ser conocido por el público en general. Esto se valoró con base en la frecuencia de aparición de la etiqueta del género por parte de los usuarios en servicios de música en línea como i-tunes.
- que contengan elementos que sean perceptualmente diferenciables. Por ejemplo, el jazz y el blues son muy parecidos por devenir uno del otro, pero la cadencia y el uso de ciertos instrumentos musicales hacen que el escucha perciba las diferencias entre ellos.

4.2 Selección de los Datos.

Se seleccionaron subconjuntos de géneros de los elegidos de acuerdo a los criterios mencionados con el fin de obtener un entendimiento de cómo se agrupaban con respecto a las características acústicas, tono o pitch (P), timbre (T) e intensidad (I). Se eligieron estas características por ser las que representan acústicamente a las melodías. Las canciones se eligieron del conjunto de datos reportados por TuWien (géneros pre-asignados por ALLMUSIC[1]).

4.2.1 Primer Conjunto de Géneros Musicales BD-1.

Se tomaron los siguientes géneros musicales: Blues, Folk, Jazz y Reggae, se generó un conjunto de 200 canciones por cada género para un total de 800 canciones. Cada canción contiene una cantidad determinada de segmentos que depende de su duración.

El pitch y el timbre están representados por una matriz M_{12xN} (ver apéndice A), donde N es el número de segmentos. La Intensidad esta representada por un vector V_{1xN}. Por lo tanto una canción contribuye con N ejemplos formados por la unión de los segmentos del $P^{(i)}$, $T^{(i)}$ e $I^{(i)}$ correspondientes al $i-segmento$ donde $i \in N$.

La distribución de ejemplos se muestra en la tabla 4.1.

Tabla 4.1: Géneros y ejemplos de la BD-1.

Género Musical	No. de Ejemplos	% Ejemplos
Blues	170699	23 %
Folk	149231	20 %
Jazz	225089	30 %
Reggae	198292	27 %
Total	743311	100 %

Distribución de los ejemplos del conjunto de 200 canciones por géneros.

[1] www-allmusic.com

4.2.1.1 **Experimento 1.**

Se aplicó K-means con las siguientes especificaciones:

- Se determinó trabajar con 4 agrupamientos o clusters,
- Se utilizó distancia euclidiana,
- Se utilizaron los cuatro primeros coeficientes del vector de timbres que están descritos en la documentación (Ver apéndice A) y
- Se evaluó utilizando el género pre-asignado en ALLMUSIC.

Se utilizó un conjunto de configuraciones para observar cómo las características seleccionadas afectaban el agrupamiento.

$$Configuraciones = \{P, T, I, PT, PI, TI, PTI\}$$

De las siete configuraciones posibles con las características I,P y T, las más relevantes fueron aquellas que consideraron las variables de forma individual. Los resultados más significativos se muestra en la tabla 4.2, los demás resultados pueden consultarse en el apéndice F. Por ejemplo, en la tabla se muestra que P logra clasificar el 33 % de los 198292 ejemplos relacionados con el género Reggae, I el 41 % y T 48 %. de forma similar para los demás géneros se indica el porcentaje de ejemplos clasificados según la etiqueta asignada previamente. La clasificación general indica el número de ejemplos clasificados correctamente del total de ejemplos.

Tabla 4.2: Resultados del experimento 1.

Género Musical	P	T	I
Reggae	33 %	48 %	41 %
Blues	27 %	31 %	38 %
Jazz	24 %	25 %	26 %
Folk	33 %	17 %	7 %
Clasificación General	28.99 %	**30.72 %**	29.09 %

Porcentaje de ejemplos agrupados correctamente por género.

4.2.1.2 **Experimento 2.**

Este experimento consistió en repetir el experimento 1 utilizando el número total de coeficientes del vector de timbres. La comparación de los resultados obtenidos se muestran en la tabla 4.3.

Tabla 4.3: Comparación de los resultados de los experimentos 1 y 2.

Género Musical	T_4	T_{12}
Reggae	48 %	51 %
Blues	31 %	32 %
Jazz	25 %	26 %
Folk	17 %	17 %
Clasificación General	30.72 %	**32.19 %**

Los sub-índices de T_4 y T_{12} indican el número de elementos del *Timbre* utilizados.

4.2.1.3 **Experimentos 3 y 4.**

Se repitieron los experimentos 1 y 2 balanceando la BD-1. La cantidad de ejemplos utilizados por cada género se limito a la cantidad correspondiente a aquel con menor número de ejemplos (en este caso Folk). La Comparación de los resultados obtenidos se muestran en la tabla 4.4.

Tabla 4.4: Comparación de los resultados de los experimentos 1, 2, 3 y 4

Género Musical	T_4	T_{12}	Tb_4	Tb_{12}
Reggae	48 %	51 %	35 %	51 %
Blues	31 %	32 %	43 %	33 %
Jazz	25 %	26 %	18 %	19 %
Folk	17 %	17 %	32 %	25 %
Clasificación General	30.72 %	**32.19 %**	32.18 %	31.82 %

Los sub-índices de Tb_4 y Tb_{12} indican el número de elementos del *Timbre* utilizados y la b de BD balanceada.

4.2.2 Segundo Conjunto de Géneros Musicales BD-2.

Se tomaron los siguientes géneros musicales: Pop Rock, Electrónica, Jazz y Rap, se generó un conjunto de 200 canciones por cada género para un total de 800. La distribución de ejemplos de la BD-2 se muestra en la tabla 4.5.

Tabla 4.5: Géneros y ejemplos de la BD-2.

Género Musical	No. de Ejemplos	% Ejemplos
Pop Rock	156312	18 %
Electrónica	298316	34 %
Jazz	225089	26 %
Rap	197589	22 %
Total	877306	100 %

Distribución de los ejemplos del conjunto de 200 canciones por géneros.

Se efectuaron los experimentos del 1 al 4 de la sección anterior utilizando ahora la BD-2. Los resultados más significativos obtenidos se muestran en las Tablas 4.6 y 4.7.

Cabe resaltar, que utilizar el total de elementos del timbre y/o balancear la base de datos, no es garantía de obtener una mejor clasificación como se muestra en las tablas 4.4 y 4.7.

Tabla 4.6: Resultados del experimento 5.

Género Musical	P	T	I
Pop Rock	22 %	47 %	5 %
Electrónica	35 %	35 %	39 %
Jazz	42 %	31 %	30 %
Rap	21 %	32 %	3 %
Clasificación General	31.07 %	**35.39 %**	31.71 %

Porcentaje de ejemplos agrupados correctamente por género.

Tabla 4.7: Comparación de los resultados de los experimentos 5, 6, 7 y 8

Género Musical	T_4	Tb_4	T_{12}	Tb_{12}
Pop Rock	47 %	42 %	5 %	58 %
Electrónica	35 %	35 %	33 %	26 %
Jazz	31 %	31 %	36 %	20 %
Rap	32 %	31 %	54 %	23 %
Clasificación General	**35.39** %	34.83 %	33.37 %	31.62 %

Los sub-índices de Tb_4 y Tb_{12} indican el número de elementos del *Timbre* utilizados y b de BD balanceada.

4.2.3 Conjuntos de Géneros Musicales BD-1m y BD-2m (Medias).

Se consideró la falta de descripción del atributo *segments_timbre* de MSD y la segunda mejor clasificación lograda para cada una de las BD's (ver tablas 4.4 y 4.7) para proponer el uso de las medias o promedios por columna de los vectores de las características como una nueva forma de representación que nos permite balancear la BD. Así se generaron BD-1m y BD-2m (Tabla 4.8).

Tabla 4.8: Géneros y No. de ejemplos. BD-1m y BD-2m.

Géneros		No. Ejemplos/género	% Ejemplos/género
Blues	Pop Rock	200	25 %
Folk	Electrónica	200	25 %
Jazz	Jazz	200	25 %
Reggae	Rap	200	25 %
Total		800	100 %

Los mejores resultados obtenidos considerando estas restricciones usando BD-1 y BD-2, se muestran en la tabla 4.9.

Tabla 4.9: Resultados obtenidos utilizando la configuración Tb_4.

Género	%ECC x Género	No. ECC
Reggae BD-1	35 %	52530
Blues BD-1	43 %	64403
Jazz BD-1	18 %	26975
Folk BD-1	32 %	48179
Pop Rock BD-2	42 %	62721
Electrónica BD-2	35 %	52428
Rap BD-2	31 %	46430
Jazz BD-2	31 %	46343

El sub-índice de Tb_4 indica el número de elementos del *Timbre* utilizados y la b de BD balanceada. ECC (Ejemplos Correctamente Clasificados). El número de ejemplos por género es de 149231 en total 596924.

4.2.3.1 **Experimentos 9-12.**

Los experimentos 9 y 10 usan BD-1m y los experimentos 11 y 12 usan BD-2m. Se realizaron de la misma forma en que se llevaron acabo los experimentos 3 y 4 o 7 y 8. Los Resultados de las configuraciones más significativas se muestran en la tablas 4.10 para BD-1m y 4.11 para BD-2m.

Tabla 4.10: Comparación de los resultados de los experimentos 9 y 10.

Género Musical	P	T_4	T_{12}	I
Folk	30 %	42 %	36 %	37 %
Jazz	32 %	26 %	44 %	20 %
Blues	26 %	39 %	27 %	33 %
Reggae	66 %	70 %	76 %	38 %
Clasificación General	38.50 %	44.13 %	**45.38 %**	31.88 %

Los sub-índices de T_4 y T_{12} indican el número de elementos del *Timbre* utilizados.

Tabla 4.11: Comparación de los resultados de los experimentos 11 y 12

Género Musical	P	T_4	T_{12}	I
Jazz	41 %	52 %	37 %	24 %
Electrónica	37 %	32 %	35 %	37 %
Rap	37 %	60 %	66 %	50 %
Pop Rock	38 %	65 %	59 %	48 %
Clasificación General	38.50 %	**52.00 %**	49.00 %	39.50 %

Los sub-índices de T_4 y T_{12} indican el número de elementos del *Timbre* utilizados.

Se observa que las medias por columna mejoran la clasificación. El comportamiento debido al uso de sólo cuatro o el total de elementos del *Timbre* no cambia respecto a los experimentos anteriores.

4.2.4 Conjunto de Géneros Musicales BD-3m.

Para esta base de datos se utilizaron algunos de los géneros ya utilizados: *Blues*, *Electrónica* y *Rap*, *Clásica* como nuevo género. Cada género con 200 canciones. Estos experimentos (13 y 14) se realizaron de la misma forma que los experimentos anteriores.

Los resultados de estos experimentos se muestran en la tabla 4.12.

Tabla 4.12: Comparación de los resultados de los experimentos 13 y 14

Género Musical	P	T_4	T_{12}	I
Blues	27 %	39 %	29 %	24 %
Rap	54 %	84 %	67 %	37 %
Clásica	60 %	52 %	47 %	50 %
Electrónica	43 %	34 %	56 %	48 %
Clasificación General	45.75 %	**51.88 %**	49.50 %	39.50 %

Los sub-índices de T_4 y T_{12} indican el número de elementos del *Timbre* utilizados.

En las tablas 4.10, 4.11 y 4.12 se observa que no necesariamente el uso de los 12 campos del *Timbre*, garantiza una mejor clasificación.

Estos resultados motivaron a utilizar sólo los primeros cuatro elementos del vector de medias del atributo *Segments_Timbre*.

Además se eliminó el género Electrónica por que el número de segmentos es mucho mayor que el de los demás géneros. Así como el género Folk por contar con el menor número de segmentos y ser un género no tan reconocido y versátil.

Para el resto del proyecto se utilizó un conjunto de datos con seis géneros: Jazz, Reggae, Blues, Rap, Clásica y Pop Rock, mismo número de géneros utilizados en la base de datos ISMIRgenre.

Se eligieron 400 canciones de forma aleatoria por género. Un total de 2400 canciones. El Conjunto se dividió en 50 % para entrenamiento, 25 % para prueba y 25 % para validación.

4.3 Análisis y Selección de Técnicas de Aprendizaje Maquinal.

Se opto por utilizar las siguientes técnicas:

- Aprendizaje No Supervisado
  - K-means por ser un algoritmo que busca agrupar a los miembros con características similares y separar lo más posibles a los grupos formados.
- Aprendizaje Supervisado
  - Perceptrón Multicapa (MLP) por formar parte de la base del desarrollo de las redes de creencia profunda (DBN).
  - Maquinas de Soporte Vectorial (SVM) por ser un clasificador más reciente y utilizado en las referencias consultadas.
- Finalmente la DBN por ser una clase de red neuronal que utiliza un entrenamiento no supervisado, ofrece características que favorecen la optimización del resultado y es utilizada en casos donde se cuenta con gran cantidad de características y datos.

En éste conjunto de experimentos se utilizó sólo la configuración Tb_4.
Para estos experimentos se utilizó el conjunto de entrenamiento (Tabla 4.13).

Tabla 4.13: Géneros Seleccionados.

Género Musical	No. de Ejemplos
Blues	200
Clásica	200
Rap	200
Pop Rock	200
Jazz	200
Reggae	200
Total	1200

4.3.1 Método no Supervisado: K-means.

Se aplicó K-means con las siguientes características:

- 6 Clusters,
- distancia euclidiana,
- vector del Timbre con los primeros 4 coeficientes y
- se evaluó utilizando el género pre-asignado a la Base de Datos con 6 géneros.

Los resultados de estos experimentos se muestran en la tabla 4.14 y en la figura 4.1.

Tabla 4.14: Matriz de Confusión con porcentajes (K-means).

Género	C0	C1	C2	C3	C4	C5	Suma
Rap	0.08 %	1.83 %	2.50 %	6.58 %	0.08 %	5.58 %	16.67 %
Blues	3.00 %	5.17 %	4.25 %	0.50 %	0.25 %	3.50 %	16.67 %
Jazz	1.08 %	3.42 %	6.92 %	2.58 %	0.58 %	2.08 %	16.67 %
Reggae	0.17 %	3.92 %	1.92 %	8.58 %	0.00 %	2.08 %	16.67 %
Clásica	4.00 %	2.58 %	4.67 %	0.42 %	4.17 %	0.83 %	16.67 %
Pop Rock	0.67 %	2.58 %	2.58 %	1.83 %	0.08 %	8.92 %	16.67 %
Suma	9.00 %	19.50 %	22.83 %	20.50 %	5.17 %	23.00 %	100 %

El porcentaje general de clasificación es de 33.83 %.

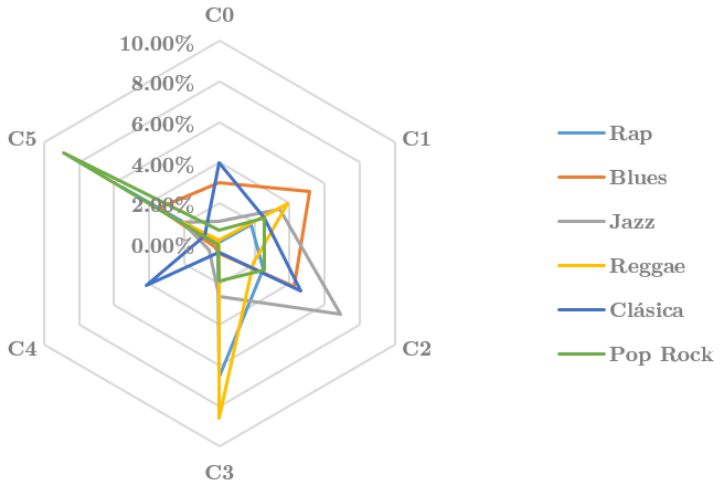

Figura 4.1: K-means. Distribución de los datos por clase.

4.3.2 Método Supervisado: Perceptrón Multicapa.

Se utilizó Perceptrón Multicapa con las siguientes características:

- 4 nodos de entrada
- 6 nodos de salida
- 5 nodos en la capa oculta
- Taza de aprendizaje de 0.3
- Momentum de 0.2
- 500 épocas
- Entrenamiento con validación cruzada de 10 pliegues
- Se utilizó el conjunto de entrenamiento de la Base de Datos con 6 géneros.

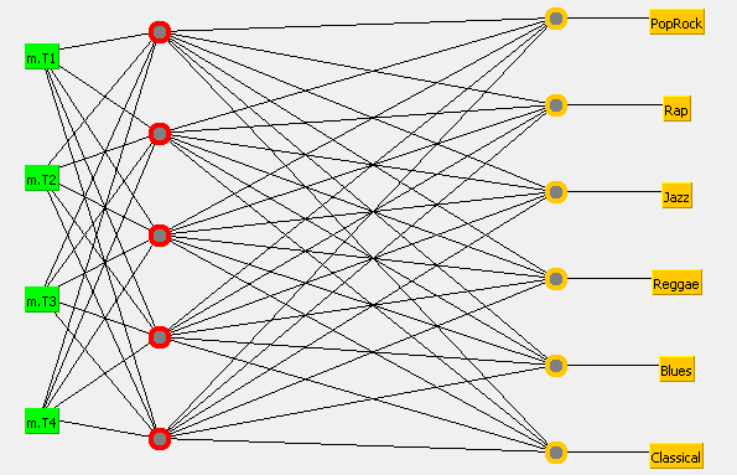

Figura 4.2: Estructura de la red neuronal.
Entradas y salidas de la MLP.

Los resultados de estos experimentos se muestran en la tabla 4.15 y en la figura 4.3.

Tabla 4.15: Matriz de Confusión con porcentajes (MLP).

Género	A	B	C	D	E	F	Suma
Pop Rock	8.50 %	1.92 %	1.42 %	1.25 %	2.50 %	1.08 %	16.67 %
Rap	1.75 %	7.17 %	1.17 %	5.08 %	0.92 %	0.58 %	16.67 %
Jazz	0.67 %	1.92 %	7.50 %	2.25 %	2.42 %	1.92 %	16.67 %
Reggae	0.75 %	4.00 %	2.00 %	9.00 %	0.83 %	0.08 %	16.67 %
Blues	3.33 %	1.00 %	3.08 %	0.92 %	6.42 %	1.92 %	16.67 %
Clásica	1.08 %	0.50 %	2.17 %	0.17 %	1.67 %	11.08 %	16.67 %
Suma	16.08 %	16.50 %	17.33 %	18.67 %	14.75 %	16.67 %	100 %

El porcentaje general de clasificación es de 49.67 %.

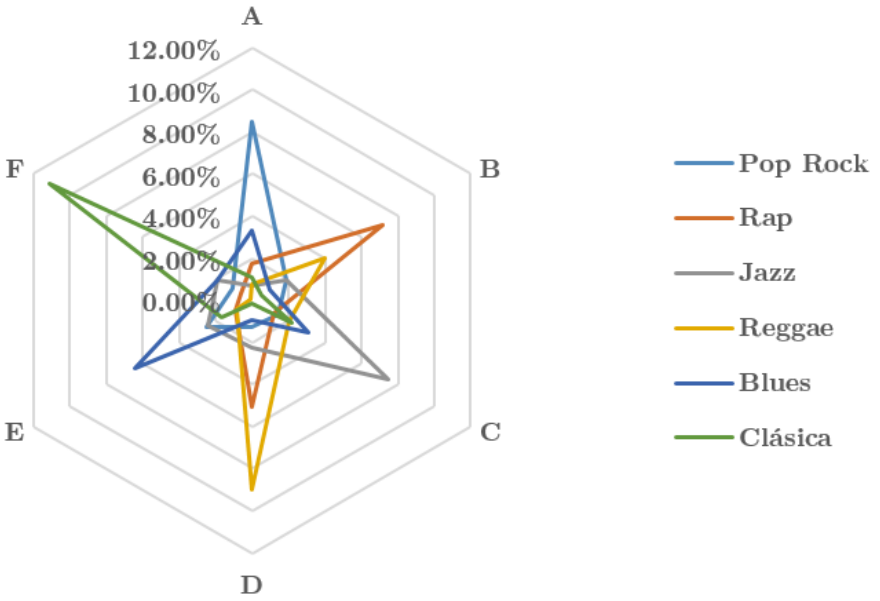

Figura 4.3: MLP. Distribución de los datos por clase.

4.3.3 Método Supervisado: Máquinas de Soporte Vectorial.

Se utilizaron Maquinas de Soporte Vectorial con diferentes tipos y configuraciones de Kernel. El mejor resultado se obtuvo con las siguientes características:

- Error de redondeo de 1.0E-12
- Datos Normalizados
- PolyKernel de orden 2
- Parámetro de tolerancia 0.001
- Entrenamiento con validación cruzada de 10 pliegues
- Se utilizó el conjunto de entrenamiento de la Base de Datos con 6 géneros

Los resultados de estos experimentos se muestran en la figura 4.4 y en la tabla 4.16.

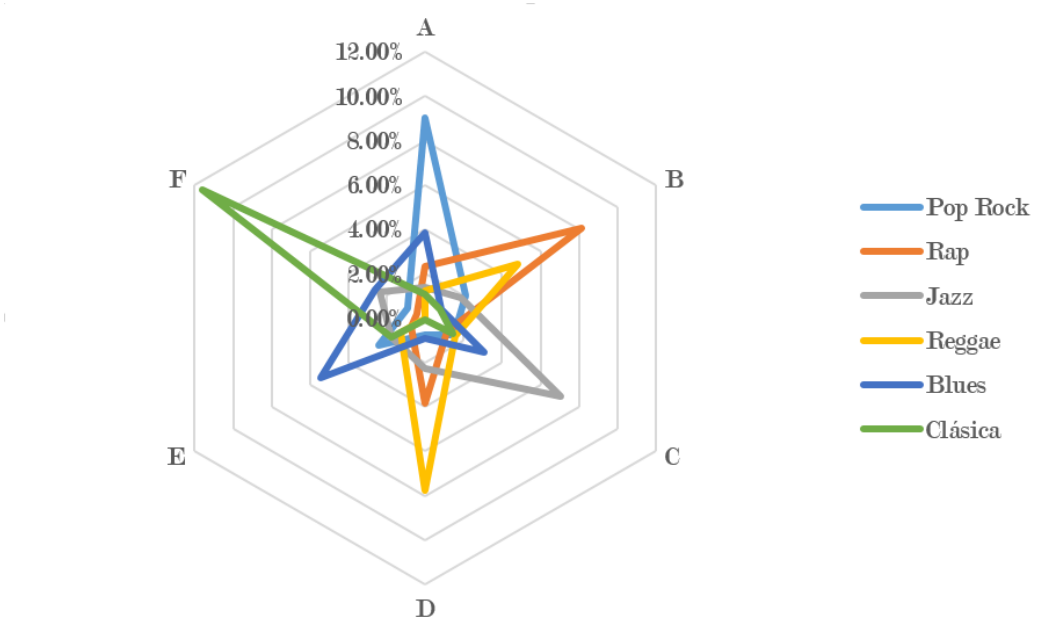

Figura 4.4: SVM. Distribución de los datos por clase.

Tabla 4.16: Matriz de Confusión con porcentajes (SVM).

Género	A	B	C	D	E	F	Suma
Pop Rock	9.00 %	2.08 %	1.50 %	0.75 %	2.42 %	0.92 %	16.67 %
Rap	2.33 %	8.17 %	1.17 %	3.83 %	0.75 %	0.42 %	16.67 %
Jazz	1.42 %	1.83 %	7.08 %	2.25 %	1.75 %	2.33 %	16.67 %
Reggae	1.25 %	4.83 %	1.58 %	7.75 %	1.25 %	0.00 %	16.67 %
Blues	3.83 %	0.83 %	3.08 %	0.92 %	5.42 %	2.58 %	16.67 %
Clásica	1.08 %	0.75 %	1.42 %	0.08 %	1.75 %	11.58 %	16.67 %
Suma	18.92 %	18.50 %	15.83 %	15.58 %	13.33 %	17.83 %	100 %

El porcentaje general de clasificación es de 49.00 %.

4.4 Pruebas, Ajustes y Evaluación.

4.4.1 Redes de Creencia Profunda (DBN).

Las redes de creencia profunda están compuestas por un conjunto de RBM's apiladas. Para obtener una visión más a fondo de qué tan bien puede trabajar la RBM con los atributos seleccionados se utilizaron distintas configuraciones y se hicieron 20 corridas por cada configuración para obtener la taza de error promedio. El máximo de iteraciones por experimento se fijo en 500. Se incorporaron las varianzas por columna de los atributos P, T e I a la base de datos. Para estos experimentos se utilizó el conjunto de entrenamiento.

En la tabla 4.17se muestra los resultados obtenidos después de entrenar una RBM. Se indica la configuración de atributos utilizada, la cantidad de nodos de entrada y salida, la tasa de error de entrenamiento mínima obtenida y la tasa de error de entrenamiento promedio. Las configuraciones que incluyen la varianza están descritas por el sub índice v.

Una vez resulta la parte de entrenar la RBN se procedió a implementar la primer DBN con una capa oculta. El número de nodos de entrada esta relacionado directamente con la configuración utilizada. El número de nodos de salida es de seis, definido por la cantidad de géneros. El número de nodos de la capa oculta es un

múltiplo de dos entre 2 y 40. Los mejores dos resultados de las seis configuraciones con mayor rendimiento se muestran en la tabla 4.18.

Tabla 4.17: RMB-1. Tasas de Error.

Configuración	Nodos X Capa	ER_{train}	$ER_{train}Prom$
PT	[24 6]	0.508333	0.553194
T	[12 6]	0.500556	0.562667
TI	[13 6]	0.505000	0.571528
P_v	[24 6]	0.593889	0.597444
PTI	[25 6]	0.497222	0.599083
P	[12 6]	0.623889	0.628250
T_4	[4 6]	0.601667	0.646500
PI_v	[48 6]	0.758333	0.815472
I_v	[2 6]	0.778333	0.823694
PI	[13 6]	0.786111	0.830306
PTI_v	[50 6]	0.833333	0.833333
PT_v	[48 6]	0.833333	0.833333
TI_v	[26 6]	0.833333	0.833333
I	[2 6]	0.833333	0.833333
T_v	[24 6]	0.833333	0.833333
T_{4v}	[8 6]	0.833333	0.833333

Ordenado por tasa de error promedio ($ER_{train}Prom$).

Tabla 4.18: DBN-1. Tasas de Error.

Configuración	Nodos X Capa	ER_{train}
P_v	[24 10 6]	0.563333
P_v	[24 8 6]	0.571667
P	[12 8 6]	0.622778
P	[12 6 6]	0.626111
T	[12 34 6]	0.660000
T	[12 28 6]	0.713333
PT	[24 38 6]	0.713333
PT	[24 24 6]	0.713889
T_4	[4 38 6]	0.713889
T_4	[4 28 6]	0.716111
PI	[13 26 6]	0.751111
PI	[13 14 6]	0.793889

Ordenado por tasa de error mínima (ER_{train}).

Se repitieron los experimentos anteriores con los cambios propuestos y los resultados obtenidos se muestran en las tablas 4.19 y 4.20.

Tabla 4.19: RMB-2. Tasas de Error(BD normalizada).

Configuración	Nodos X Capa	ER_{train}	$ER_{train}Prom$
PTI_v	[50 6]	0.338333	0.342972
PT_v	[48 6]	0.342222	0.346944
TI_v	[26 6]	0.357222	0.362861
T_v	[24 6]	0.358889	0.363250
PTI	[25 6]	0.394444	0.399028
PT	[24 6]	0.397778	0.400028
TI	[13 6]	0.428333	0.431361
T	[12 6]	0.430556	0.433611
T_{4v}	[8 6]	0.481111	0.482556
T_4	[4 6]	0.549444	0.552556
PI_v	[26 6]	0.546667	0.552639
P_v	[24 6]	0.594444	0.597972
PI	[13 6]	0.595000	0.599722
P	[12 6]	0.625000	0.628222
I_v	[2 6]	0.684444	0.688278
I	[1 6]	0.707222	0.712417

Ordenado por tasa de error promedio ($ER_{train}Prom$).

Tabla 4.20: DBN-2. Tasas de Error (BD normalizada).

Configuración	Nodos X Capa	ER_{train}
PTI_v	[50 20 6]	0.265000
PTI_v	[50 18 6]	0.273889
PT_v	[48 18 6]	0.278889
PT_v	[48 22 6]	0.282778
TI_v	[26 10 6]	0.317778
TI_v	[26 12 6]	0.341111
T_v	[24 12 6]	0.339444
T_v	[24 6 6]	0.350000
PT	[24 12 6]	0.360000
PT	[24 18 6]	0.364444
PTI	[25 14 6]	0.363333
PTI	[25 16 6]	0.365556
T	[12 8 6]	0.416667
T	[12 6 6]	0.422222
TI	[13 8 6]	0.417222
TI	[13 6 6]	0.423889
PI_v	[26 10 6]	0.499444
PI_v	[26 12 6]	0.520000
T_{4v}	[8 4 6]	0.526111
T_{4v}	[8 2 6]	0.575000
P_v	[24 8 6]	0.554444
P_v	[24 10 6]	0.560556
PI	[13 12 6]	0.558333
PI	[13 10 6]	0.561667
P	[12 8 6]	0.599444
P	[12 10 6]	0.610556
T_4	[4 2 6]	0.613333
T_4	[4 4 6]	0.825556

Ordenado por tasa de error mínima (ER_{train}).

Finalmente se entreno la red utilizando la mejor configuración de atributos lograda, y se efectuaron la validación y prueba de la red utilizando los conjuntos correspondientes. La matriz de confusión para los conjuntos de entrenamiento, validación y prueba se muestra en las tablas 4.21, 4.23 y 4.22 respectivamente.

Tabla 4.21: Matriz de Confusión con porcentajes (Train DBN).

Género	Reggae	Rap	Pop Rock	Jazz	Clásica	Blues	Suma
Reggae	12.11 %	2.22 %	0.83 %	0.22 %	0.39 %	0.89 %	16.67 %
Rap	2.28 %	12.67 %	0.72 %	0.28 %	0.22 %	0.50 %	16.67 %
Pop Rock	0.78 %	0.72 %	12.22 %	0.50 %	0.72 %	1.72 %	16.67 %
Jazz	1.50 %	0.28 %	0.83 %	10.50 %	1.33 %	2.22 %	16.67 %
Clásica	0.28 %	0.39 %	0.78 %	0.56 %	13.56 %	1.11 %	16.67 %
Blues	1.00 %	0.22 %	1.83 %	1.17 %	0.44 %	12.00 %	16.67 %
Suma	17.94 %	16.50 %	17.22 %	13.22 %	16.67 %	18.44 %	100 %

El porcentaje general de clasificación es de 73.06 %.

Tabla 4.22: Matriz de Confusión con porcentajes (Validation DBN).

Género	Reggae	Rap	Pop Rock	Jazz	Clásica	Blues	Suma
Reggae	10.67 %	4.00 %	0.00 %	0.50 %	0.17 %	1.33 %	16.67 %
Rap	1.83 %	12.83 %	1.00 %	0.33 %	0.33 %	0.33 %	16.67 %
Pop Rock	0.67 %	0.83 %	11.50 %	0.50 %	1.50 %	1.67 %	16.67 %
Jazz	1.33 %	0.33 %	1.83 %	7.00 %	3.00 %	3.17 %	16.67 %
Clásica	0.00 %	0.67 %	1.50 %	0.50 %	13.33 %	0.67 %	16.67 %
Blues	2.33 %	0.33 %	2.67 %	1.00 %	0.50 %	9.83 %	16.67 %
Suma	16.83 %	19.00 %	18.50 %	9.83 %	18.83 %	17.00 %	100.00 %

El porcentaje general de clasificación es de 65.17 %.

Tabla 4.23: Matriz de Confusión con porcentajes (Test DBN).

Género	Reggae	Rap	Pop Rock	Jazz	Clásica	Blues	Suma
Reggae	10.33 %	2.83 %	0.00 %	0.67 %	0.17 %	2.67 %	16.67 %
Rap	3.00 %	11.17 %	1.33 %	0.00 %	0.33 %	0.83 %	16.67 %
Pop Rock	0.50 %	0.67 %	11.17 %	1.67 %	1.50 %	1.17 %	16.67 %
Jazz	1.33 %	0.00 %	1.50 %	10.00 %	1.83 %	2.00 %	16.67 %
Clásica	0.00 %	0.33 %	0.50 %	1.00 %	13.50 %	1.33 %	16.67 %
Blues	1.50 %	0.33 %	2.00 %	1.00 %	0.67 %	11.17 %	16.67 %
Suma	16.67 %	15.33 %	16.50 %	14.33 %	18.00 %	19.17 %	100.00 %

El porcentaje general de clasificación es de 67.33 %.

The important thing is not to stop questioning. Curiosity has its own reason for existing.

Albert Einsten

CAPÍTULO

5

Discusión

El reconocimiento de géneros musicales puede verse como un problema de reconocimiento de patrones donde el objetivo es definir los patrones que puedan diferenciar cada clase. Durante la investigación se encontraron con los siguientes aspectos relevantes.

5.1 Elección de Géneros Musicales.

La definición de género musical se vuelve más compleja al surgir nuevos géneros como consecuencia de la fusión de los ya existentes. Los géneros utilizados en BD-1 fueron seleccionados de acuerdo a lo reportado en la literatura con respecto al diseño e implementación de reconocedores. En BD-2 y BD-3m se optó por combinar los géneros que cumplieran con los criterios descritos y que tuvieran suficientes ejemplos para realizar los experimentos.

La inclusión del género folk no resultó positiva por ser muy dependiente de su contexto inmediato (no es tan conocido en México como lo es en EEUU); el rap por su parte, aunque también puede considerársele como un género contextualizado, su generalización ha sido más aceptada en diferentes lugares geográficos por lo que

se decidió incorporarla al conjunto final. En este sentido consideramos que un sistema de reconocimiento de géneros musicales deberá incorporar de alguna manera aspectos no medibles directamente en el audio a fin de considerar géneros como estos. La MSD cuenta con información no numérica que podría ser utilizada para este fin.

El conjunto final de géneros incluye algunos que aparecen en la base de datos GTZAN y la misma cantidad de ISMIRgenre: Blues, Clásica, Jazz, Pop Rock, Rap y Reggae. Si bien estos géneros poseen una amplia difusión global, un sistema robusto de reconocimiento de géneros musicales tendría que considerar un mayor número de éstos dada la dinámica creativa en constante evolución del mundo de la música.

5.2 Elección de Coeficientes del Timbre.

En un principio se eligieron como características aquellas que describieran acústicamente las canciones de cada género, éstas fueron pitch o tono (P), timbre (T) y sonoridad o intensidad (I). Al realizar los experimentos se encontró que los mejores porcentajes de ejemplos clasificados correctamente (%ECC) correspondían a las configuraciones que incluían T.

Al experimentar con la BD-1 (experimentos 1-4) se encontró lo siguiente:

- la configuración P muestra los menores %ECC
- el uso del timbre ayuda a mejorar la clasificación
- el balanceo de los datos mejora la clasificación con respecto a las otras configuraciones
- el balanceo de los datos y el uso de los primeros cuatro coeficientes T mejora ligeramente la clasificación con respecto a usar el balanceo de datos en el total de coeficientes T, pero ligeramente inferior a usar todos los coeficientes T sin balanceo de datos.

De acuerdo a los resultados podemos concluir que:

$$\%ECC_{T_4} < \%ECC_{Tb_{12}} < \%ECC_{Tb_4} < \%ECC_{T_{12}}$$

Por su parte, al experimentar con la BD-2 (experimentos 5-8) se encontró lo siguiente:

- los coeficientes T ofrecen los menores %ECC
- la incorporación de los coeficientes T no garantiza una mejora en el %ECC
- el balanceo de los datos no garantiza una mejora en el %ECC
- el balanceo de los datos con el total de coeficientes T no mejora los resultados con respecto al uso de los primeros 4 coeficientes T

De acuerdo a los resultados podemos concluir que

$$\%ECC_{Tb_{12}} < \%ECC_{T_{12}} < \%ECC_{Tb_4} < \%ECC_{T_4}$$

En este punto se observó que las características del timbre ofrecían menores resultados aunque la configuración de sus coeficientes no coincidió en todos los experimentos hasta ahora desarrollados. En este sentido se decidió tomar como vector de entrada los primeros cuatro coeficientes de los datos balanceados (Tb_4) que fue la que con mayor frecuencia apareció en los mejores dos resultados por experimento.

Al realizar los experimentos con las BDm-1 y BDm-2 (experimentos 9-10 y 11-12) se encontró que

$$\%ECC_{T_4} < \%ECC_{T_{12}}$$
$$\%ECC_{T_{12}} < \%ECC_{T_4}$$

con la BDm-3 se obtuvo lo siguiente

$$\%ECC_{T_{12}} < \%ECC_{T_4}$$

Al final se decidió utilizar como vector de características el compuesto por los 4 primeros coeficientes de T dado que la diferencia entre éste y el vector de 12 coeficientes de T es mínima y que en la documentación sólo se describen a estos cuatro coeficientes.

5.3 Balanceo de los Datos.

Las bases de datos que se conformaron para los experimentos no estaban balanceadas, es decir, no tenían el mismo número de canciones por género ni el mismo número de segmentos por canción. Esto ocasionaba que se pudiera presentar una clase dominante por el número de instancias. Para solucionar el primer problema se tomó el mismo número de canciones por género y para el segundo se tomó el número de segmentos igual al menor número de los mismos de entre los géneros considerados para la BD, en este caso dicho género fue el folk. Aún con estas consideraciones se deben tomar en cuenta aspectos como el punto de inicio para seleccionar los segmentos de las canciones que pertenecen a los géneros con mayor número de éstos. En el caso del género folk, éste contenía 149231 segmentos; para obtener este número de segmentos en los otros géneros se procedió desde el principio del conjunto de canciones por género, es decir, se tomaron los primeros 149231 segmentos de los 170699 de blues y así sucesivamente para los demás (figura 5.1). Una opción sería tomar los últimos 149231 segmentos o centrarlo, es decir, tomar del segmento 10734 al 159964 de blues.

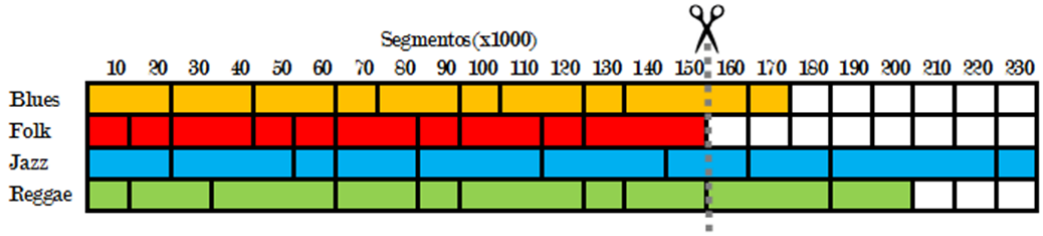

Figura 5.1: Balanceo por segmentos.

El balanceo por segmentos implica la reducción y/o eliminación de canciones.

5.4 Evaluación de la Capacidad de Agrupamiento.

En la metodología establecida se procedió a utilizar una técnica de aprendizaje por agrupamientos para definir el conjunto de datos que se ocuparían en cuanto a las clases (géneros musicales) y atributos (características). Se partió de la premisa de que, debido a que en teoría un conjunto de canciones del mismo género musical presenta características de armonía y ritmo similares, al aplicar un agrupamiento

se esperaría que la mayoría de éstas tendiera a irse al mismo conjunto. De manera similar, los atributos finales se obtuvieron de la capacidad de agrupamiento de los coeficientes P, T e I en cada género.

La base de datos con la que se trabajó tenía un número delimitado de características acústicas y éstas se presentaban de una manera muy específica; el agrupamiento surgió como una primera aproximación para el análisis de los datos y selección de atributos para la clasificación. Esta estrategia puede modificarse si en lugar de metadatos se contara con la información de los audios de las canciones, algo que resulta difícil de conseguir por asuntos de propiedad intelectual.

5.5 Aplicación de Medias.

Balancear las bases de datos respecto al número de ejemplos, logró mejorar el %ECC al utilizar P o T pero no para I con BD-1. Sin embargo en BD-2 sucedió todo lo contrario.

Se incrementó el %ECC de los géneros que antes contaban con menor cantidad de ejemplos y para el resto el %ECC se redujo. Este balanceo benefició a la BD-1 en la mayoría de sus configuraciones pero perjudicó a la mayoría de BD-2.

Al hacer los experimentos con las bases de datos balanceadas en número de canciones y ejemplos (uso de medias) los resultados mejoraron considerablemente manteniendo el mismo comportamiento para cada uno de los elementos del conjunto de configuraciones. Sin embargo, para DBm-1 el %ECC del Blues se redujo y para DBm-2 Electrónica no fue afectado considerablemente.

Analizando la información que representan los cuatro primeros coeficientes del Timbre tenemos:

- T1: Sonoridad[1]: Cualidad de la sensación auditiva que permite apreciar la mayor o menor intensidad de los sonidos. Se mide en fonios.

[1] Definición de la RAE.

- T2: Brillo:Riqueza en sonidos agudos.
- T3: Plano: Relacionado con la dinámica y continuidad del sonido.
- T4: Ataque: Producir un sonido por medio de un golpe seco y fuerte. Por ejemplo el pizzicato en el violín.

El timbre como tal nos permite distinguir entre instrumentos o personas. Como se mencionó en el apéndice E es también conocido como el color de la música. Si pensamos en una imagen, los cuatro coeficientes que representan el Timbre se podrían interpretar como la cantidad de colores, el brillo, el degradado y el contraste.

Si pensamos en un tambor, la sonoridad será la profundidad del sonido y su grado de reverberación (gama de colores), el brillo tiene que ver con la riqueza en agudos del sonido. Lo plano esta relacionado con la falta de matices, variaciones o texturas (Liso-arrugado, degradado) y el ataque esta relacionado con el tiempo que transcurre en emitirse un sonido (el contraste). La Campana tiene un ataque mucho mayor al roce del arco con el violín.

Así el primer coeficiente podría interpretarse como una medida de la gama de frecuencias involucradas (sonoridad), el segundo coeficiente se podría interpretar como una medida de frecuencias altas, el tercer coeficiente como una medida de que tan plano es el sonido y el cuarto como una medida de lo intempestiva de la interpretación.

Con esto en mente al promediar los segmentos por columna, identificamos los puntos de equilibrio o ruptura de cada coeficiente que representan el timbre.

La mayoría de las canciones tienen inicios, partes intermedias y finales, que son diferentes al género. Dado que se están utilizando los segmentos de canciones completas, al promediar, los datos con valores atípicos relacionados con estas partes de la canciones se reducen.

5.6 Comportamiento de los Géneros.

Al observar los resultados después de aplicar aprendizaje maquinal a los datos se encontró en general que los que se agrupaban correspondían a más de un género; sin embargo, se presentaron los siguientes casos particulares:

- Datos de canciones pertenecientes a los géneros reggae y pop rock tendían a formar un agrupamiento para sí; lo mismo pasó para el jazz pero en menor medida (figura 4.1).
- En el caso del MLP así como para el de SVM los géneros clásico y pop rock fueron los que agruparon la mayoría de las canciones que pertenecían a esas clases.

Desde una perspectiva de apreciación auditiva estos géneros son relativamente fáciles de identificar debido a ciertos elementos musicales asociados a estos géneros.

5.7 Uso de Aprendizaje Profundo.

El uso del aprendizaje profundo no produjo resultados positivos en la clasificación de los géneros utilizando T_4. Se experimentó con una serie de configuraciones de redes neuronales para identificar el efecto de las mismas (número de capas, número de nodos por capa incluyendo la de salida y número de iteraciones) encontrando que no había cambio significativo en los resultados. Una posible razón de este comportamiento apuntaría a la naturaleza de los datos: como entrada se tuvo un vector de cuatro coeficientes, los cuales vienen de un proceso de información al cual no tenemos acceso. Es probable que el número reducido de coeficientes no haya podido aportar suficientes elementos para que el algoritmo pudiera encontrar una salida satisfactoria; en este sentido se tuvo que probar con otras de las configuraciones de coeficientes incorporando P o I.

Hasta este punto se utilizaron los datos que presentaron mejores resultados en las etapas previas y se asumió que éstos darían mejores respuestas.

Sin embargo al observar los resultados de la tabla 4.17 donde se incluyó los resultados de los experimentos que incluyeron la varianza relacionada con las medias de los atributos, se observó que el uso de las medias de los coeficientes del pitches junto sus varianzas, produce mejores resultados que los logrados con la configuración de características propuesta T_4. Además la configuración PTI obtuvo la menor taza de error de entrenamiento.

Los resultados obtenidos para la configuración P_v en la primer DBN confirmó lo observado y motivó a normalizar las medias y varianzas de los coeficientes del timbre y sonoridad, con base en el hecho de que P es un atributo normalizado y que el uso de las medias y varianzas de sus coeficientes mejoró su rendimiento. Por lo tanto se optó por normalizar los valores del timbre y sonoridad. Se pensó en al menos dos formas de hacerlo: normalizando los datos existentes de la BD o crear una nueva BD normalizando antes los coeficientes del timbre y después obtener las medias y varianzas de éstos. Por cuestiones de tiempo se eligió la primer opción.

Los resultados obtenidos con estos cambios son satisfactorios y muestran una tendencia al uso de la mayor cantidad posible de nodos de entrada mientras se considere el timbre como atributo principal.

De acuerdo a los resultados obtenidos el sistema que mejor clasifica los géneros musicales estudiados esta compuesto por una red con 50 nodos de entrada que corresponden con la configuración PTI_v, 24 valores relacionados con el pitch, 24 con el timbre y dos con la sonoridad. La capa oculta contiene 20 neuronas y la salida cuenta con 6 neuronas que corresponden al número de géneros utilizados, como se muestra en la figura 5.2.

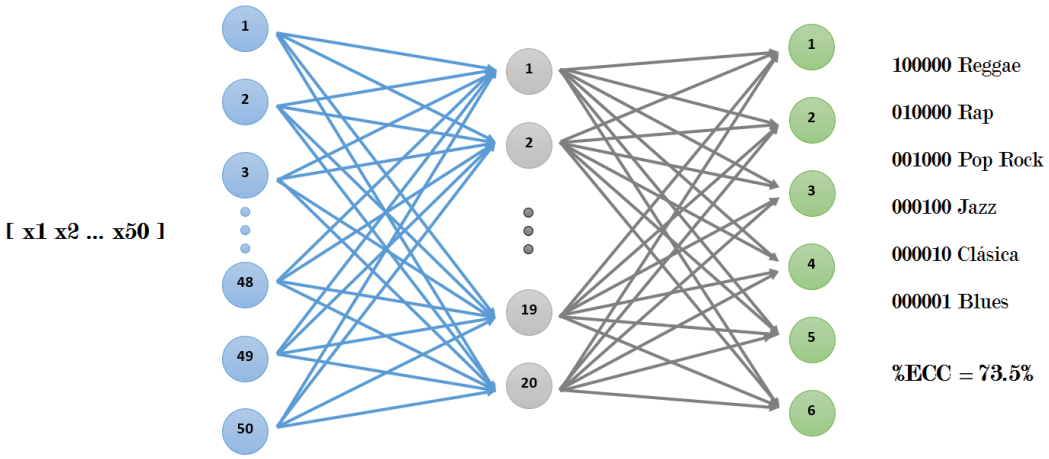

Figura 5.2: DBN. Arquitectura de la red que ofrecio los mejores resultados.

Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.

Winston Churchill

CAPÍTULO

6

Conclusiones y Trabajo Futuro

De manera general podemos decir que se cubrieron los objetivos con sus bemoles.

- Se logró desarrollar un sistema de reconocimiento de géneros musicales aplicando redes de creencia profunda, el cual no requirió más de una capa oculta. Esto debido a que la RBM pudo modelar bastante bien los datos presentados.
- Se seleccionaron los componentes que permitieron el análisis de géneros de piezas musicales.
- Se definió la arquitectura de aprendizaje maquinal para el reconocimiento y se implementó.
- Se realizó la etapa de pruebas del sistema de reconocimiento para ajustar los parámetros que lo requirieran.
- Se realizó el ajustes del sistema de reconocimiento, se realizaron pruebas finales y se evaluó el rendimiento del sistema de reconocimiento.

6. CONCLUSIONES Y TRABAJO FUTURO

En cuanto a lo particular, como resultado de los experimentos realizados en las diferentes etapas de la investigación surgieron varios puntos que caben destacar.

En la etapa de elección de los géneros musicales podemos concluir que el conjunto de géneros seleccionados es un buen conjunto pues cuenta con géneros que se han utilizado en otras bases de datos y la cantidad de ejemplos con las que cuenta es mayor. Son géneros globales, bien conocidos y reconocibles por la comunidad.

En la etapa de elección de coeficientes del timbre se efectuaron una serie de experimentos utilizando un procedimiento de aprendizaje no supervisado (K-means) con los atributos pitch, timbre y sonoridad que nos permitió concluir que el atributo más importante fue el timbre independientemente del número de coeficientes utilizado. Además se observó que las SVM obtuvieron los mejores porcentajes de ejemplos correctamente clasificados. Una hipótesis es que ésto se debió a que el código para las SVM normaliza los datos de entrada evitando de esta forma el sesgo que puede ocasionar la disparidad de los rangos de los datos del timbre y sonoridad.

En la etapa de balanceo de los datos la decisión de utilizar los atributos pitch, timbre y sonoridad para la canción completa tuvo gran impacto afectando toda la investigación. Una opción era considerar sólo algunos segundos de las partes principales de la canción, como se hace en la literatura, de esta forma podríamos haber obtenido otro conjunto de datos bien balanceado. Sin embargo el tener desbalanceada la base de datos nos mostró que no necesariamente contar con una mayor cantidad de ejemplos implica una mejor clasificación y tampoco contar con una cantidad menor de ejemplos implica una peor clasificación. Así que la selección y representación de los atributos cobró un papel muy importante.

El balanceo de la BD por cantidad de segmentos implico un desbalance en el número de canciones por género y en la mutilación de otras sin lograr cambios significativos.

En la etapa de evaluación de la capacidad de agrupamiento se determinó utilizar las características P,T e I relacionadas con la armonía y ritmo. Sin embargo

podríamos haber optado por incluir algunos metadatos como el tempo y el compás que se relacionan con el ritmo. Así que no necesariamente los atributos seleccionados fueron los mejores para esta tarea de reconocimiento de géneros musicales.

En la etapa de aplicación de medias se buscó solucionar los problemas de segmentación y/o pérdida de canciones calculando las medias del total de segmentos de la canción (P,T e I) por columnas. Los resultados obtenidos avalan que fue una buena medida que además propicio la aceleración del proceso de clasificación, pues se redujo el número de ejemplos (de cientos de miles a cientos) y se incrementó el porcentaje de ejemplos clasificados correctamente.

En la etapa de aprendizaje profundo, se tuvo problemas con la representación y la cantidad de los atributos seleccionados que se vio reflejado en el pobre rendimiento del sistema. Las DBN son una opción para cuando los datos están representados por un grupo grande de atributos y se cuenta con pocos ejemplos.

La decisión de reconsiderar la representación de los datos para realizar nuevamente los experimentos nos permitió observar la importancia que tiene la normalización de los datos. Al proyectar los datos a un intervalo entre 0 y 1 los sesgos debidos a la variabilidad de los intervalos de cada uno de los coeficientes se redujeron, razón por la cual se logró que el sistema obtuviera mejores resultados. Además de validar la hipótesis sobre la normalización de los datos en las máquinas de soporte vectorial, el pre-procesamento de normalización fue la clave para mejorar los resultados y poder incluir más características.

La tabla 6.1 muestra los resultados de trabajos relacionados con el reconocimiento de géneros musicales sobre la base de datos LM expuesto en el ISMIR y evaluados por MIREX, y Hamel con el fin de poner nuestros resultados en contexto.

Tabla 6.1: Comparativa de Trabajos Relacionados con SRGM.

Autor	Clasificador	Representación	Base de Datos	Rendimiento (%)
Costa (2012)*	SVM	3 Espectrogramas (Texturas GLCM y LBP) Zonificación n = 3,5,10,15,20	Latin Music Database	70.78 (GLCM n=5) 80.33 (LBP n=5)
Costa (2011)	SVM Esquema de mayoría de votos	3 Espectrogramas (Texturas GLCM) Zonificación n=10	Latin Music Database	67.5 (GLCM n=10)
*MIREX 2010	SVM	- - - - -	Latin Music Database	79.8
*MIREX 2009	SVM	- - - - -	Latin Music Database	74.6
*MIREX 2008	SVM	- - - - -	Latin Music Database	65.1
Hamel(2010)	SVM	DBN Activation like input	Tzanetakis	84.3
SRGM(2016)	DBN	Medias y varianzas por columnas de los coeficientes P,T e I	MSD	73.5

Rendimiento de Sistemas de Reconocimiento de Géneros Musicales.

En la tabla podemos observar que el SRGM que se logró no ofrece el mejor rendimiento, sin embargo tampoco es el peor. Las diferencias claramente están relacionadas con los datos utilizados, su representación y el método de clasificación.

MSD ofrece un conjunto de datos sobre canciones ya analizadas utilizando la API de EchoNest sin la asignación de género mientras que LMD ofrece audio de canciones con el género asignado por expertos de la música. Los géneros utilizados en este trabajo son diferentes a los contenidos en LMD y Tzanetakis. Además el conjunto de segmentos de MSD son obtenidos sobre la canción completa mientras que las otras bases ocupan solo algunas partes de la canción.

6.1 Trabajo futuro.

Este sistema se puede mejorar si se toman en cuenta los aspectos relacionados con los datos y sus procesamientos. En particular se puede

- Utilizar otras bases de datos que incluyan el archivo de audio para probar con otras técnicas aplicadas directamente a la señal cruda.
- En el caso de MSD probar con algún tipo de procesamiento del dato como la Transformada de Fourier a los vectores T de 4 y 12 coeficientes con fines de comparación con lo hecho hasta ahora.
- Alinear los vectores T de 4 y 12 coeficientes al beat y obtener las medias para la clasificación, esto con la finalidad de aglutinar y en consecuencia enfatizar los rasgos de cada género para así lograr una mayor separación.
- Aumentar la cantidad de atributos seleccionados de la MSD relacionados con las cualidades acústicas para complementar la Base de Datos utilizada con el fin de observar el rendimiento de la DBN.
- Con el fin de eliminar algunas de las similitudes entre todos los géneros se podría evitar utilizar los segmentos de secciones como el fade-in y fade-out de la canción y llevar acabo las medias para los segmentos que se encuentren en una cierta cantidad de tiempo (e.j. 30s después de finalizada la parte inicial, intermedio y antes de iniciar la parte final).
- Utilizar el conjunto de etiquetas como atributos y crear un vector del conjunto de etiquetas con sus relativos pesos relacionadas con la canción con el fin de tener un vector lo suficientemente grande que implique un reto para la DBN.

Bibliografía

[1] J. TAGUE-SUTCLIFFE, "The pragmatics of information retrieval experimentation", revisado en Information Processing and Management, vol. 28, no. 4, pp. 467-490, 1992.

[2] F. HOLT, *Genre in Popular Music*, Chicago, IL: University of Chicago Press, 2007.

[3] J. C. LENA, R. A. PETERSON, "Classification as culture: Types and trajectories of music genres", American Sociological Review, vol. 73, no. 5, pp. 697-718, October 2008.

[4] M. A. DUVAL, S. VEGA-PONS, J. RUIZ, "Combinación de clasificadores supervisados: estado del arte", Centro de aplicaciones de tecnologías de avanzada (CENATAV), La Habana Cuba, Reporte Técnico 048. Abril 2012.

[5] D. PYLE, "Data preparation for data mining, Morgan Kaufmann", Sn Fco. California, 1999.

[6] I. GUYON, A. ELISSEEFF, "An introduction to variable and feature selection", Journal of machine learning research 3, 2003, pp. 1157-1182.

[7] M. GOTO, "Scene description project: Toward audio-based real-time music understanding", en Proceedings of the 3rd International Conference on Music Information Retrieval, pp. 231–232, 2003.

[8] M. GOTO, "A real-time music scene description system: Predominant F0 estimation for detecting melody and bass lines in real-world audio signals", Speech Communication, vol. 43, no. 4, pp. 311–329, 2004.

[9] A. KLAPURI, M. DAVY, "Signal Processing Methods for Music Transcription", Springer, New York, NY, 2006.

[10] M. MÜLLER, D.P.W. ELLIS, A. KLAPURI, G. RICHARD, "Signal processing for music analysis", IEEE Journal on Selected Topics in Signal Processing, vol. 5, no. 6, pp. 1088-1110, Octubre 2011.

[11] E. HUMPHREY, J. P. BELLO, Y. LECUN, "Moving Beyond Feature Design: Deep Architectures and Automatic Learning in Music Informatics", Proceedings of 13th, ISMIR 2012, pp. 403-408, 2012.

[12] T. BERTIN-MAHIEUX, D. P. W. ELLIS, B. WHITMAN, AND P. LAMERE, "The million song dataset", en Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR), pp. 591-596, 2011.

[13] R. HILLEWAERE, B. MANDERICK, D. CONKLIN, "String Methods for Folk Tune Genre Classification", en Proceedings of the 13th International Society for Music Information Retrieval Conference, Porto, Portugal, pp. 8-12, Octubre 2012.

[14] P. HAMEL, D. ECK, "Learning features from music audio with deep belief networks", en 11th International Society for Music Information retrieval Conference ISMIR, pp. 339-344, 2010.

[15] G. E. HINTON, S. OSINDERO, Y. TEH, "A fast learning algorithm for deep belief nets", Neural Computation, vol. 18, no.7, pp. 1527-1554, Julio 2006.

[16] E. J. HUMPHREY, J. P. BELLO, AND Y. LECUN, "Deep architectures and automatic feature learning in music informatics", en ISMIR, pp. 403–408, 2012.

[17] Y. M. G. COSTA, L. S. OLIVEIRA, A. L. KOERICH, F. GOUYON, "Music genre recognition based on visual features with dynamic ensemble of classifiers selection", en 20th International Conference on Systems, Signals and Image Processing, Bucharest, Romania, IEEE Press, pp. 55-58, 2013.

[18] T. LIDY, A.RAUBER, "Evaluation of feature extractors and psychoacoustic transformations for music genre classification", en Proceedings of the 6th International Conference on Music Information Retrieval. London, UK. September 11-15, pp. 34-41, 2005.

[19] Y. M. G. COSTA, L. S. OLIVEIRA, A. L. KOERICH, F. GOUYON, "Music genre recognition using spectrograms", en 18th International Conference on Systems, Signals and Image Processing, pp. 151–154, 2011.

[20] Y. M. G. COSTA, L. S. OLIVIERA, A. L. KOERICH, F. GOUYON, J. MARTINIS, "Music genre classification using LBP textural features", en Signal Processing, vol. 92, pp. 2723-2737, 2012.

[21] Y. BENGIO, "Learning Deep Architectures for AI", Foundations and Trends in Machine Learning, vol. 2, no.1, pp. 1-127, 2009.

[22] Million Song Dataset Benchmarks, Viena University of Technology, Information and Software Engineering Group, disponible en: http:// www.ifs.tuwien.ac.at/mir/msd/partitions/msd-MAGD-genreAssignment.cls

[23] Y. BENGIO, P. LAMBLIN, D. POPOVICI, AND H. LAROCHELLE, "Greedy Layer-Wise Training of Deep Networks", en Advances in Neural Information Processing Systems, vol. 19, pp. 153-160, 2007.

[24] G. E. HINTON, R. R. SALAKHUTDINOV, "Reducing the Dimensionality of Data with Neural Networks", Science, vol. 313. no. 5786, pp. 504-507, 2006.

[25] G. E. HINTON, T. J. SEJNOWSKI,"Optimal perceptual inference", Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 448–453, 1983.

APÉNDICE

A

Descripción y Estructura de Million Song Dataset

A.1 Million Song Dataset (MSD).

La Base de datos MSD es una colección de características de audio y metadatos disponible de forma libre para un millón de pistas de música popular contemporánea proporcionada por Echo Nest. No incluye ningún audio, sólo las características derivadas. Sin embargo, el audio de la muestra puede ser recuperado de servicios como 7digital, utilizando el código que se proporciona[1].

La MSD consta también de un grupo de conjuntos de datos complementarios aportados por la comunidad:

- SecondHandSongs BD (Covers)
- musiXmatch BD (Letras)
- Last.fm BD (Etiquetas a nivel de canción y semejanza)

[1] https://github.com/tbertinmahieux/MSongsDB/tree/master/Tasks Demos/Preview7digitaltext

- Taste Profile subconjunto (Datos de usuario)

La MSD comenzó como un proyecto de colaboración entre The Echo Nest y LabROSA. Fue apoyado en parte por la NSF (National Science Foundation).

Sus objetivos son:

- Fomentar la investigación en algoritmos que se escalan a tamaños comerciales.
- Proporcionar un conjunto de datos de referencia para la evaluación de la investigación.
- Ser una alternativa de acceso libre a la creación de una gran base de datos con las API (por ejemplo, The Echo Nest's API).
- Ayudar a los nuevos investigadores a comenzar en el campo MIR.

Para descargar la MSD se requiere descargar 26 archivos comprimidos que dan un tamaño total de 266 GB (285,748,207,616 bytes) en disco una vez descomprimidos.

A.2 Organización de MSD.

La MSD consta casi de toda la información disponible para un millón de canciones obtenida a través de la API de Echo Nest. Esto incluye tanto los metadatos como las características del análisis de audio. Existe un archivo para cada pista que corresponde a una canción, un lanzamiento y una artista. Toda la información sobre estos cuatro elementos: pista, canción, lanzamiento y artista, está en todos los archivos lo que implica una cierta redundancia, aunque la mayor parte de los datos relacionados con el análisis de audio, son únicos. En la figura A.1 se muestra un esquema donde se puede observar la forma en que se lleva a cabo la organización interna de los atributos del archivo de la canción en formato HDF5.

A.3 Estructura y Descripción de MSD.

A.3.1 Los Valores de Confianza (Confidence)

Muchos elementos de la pista y del análisis incluyen valores de confianza, un número de punto flotante que va de 0.0 a 1.0. Confidence indica la fiabilidad o confianza de su atributo correspondiente. Los elementos que llevan un pequeño valor de confianza deben considerarse especulativos. Puede que no haya datos suficientes en el audio para calcular el elemento con gran certeza. Se almacenan en un arreglo de punto flotante de 64 bits cuya longitud depende de la duración de la canción.

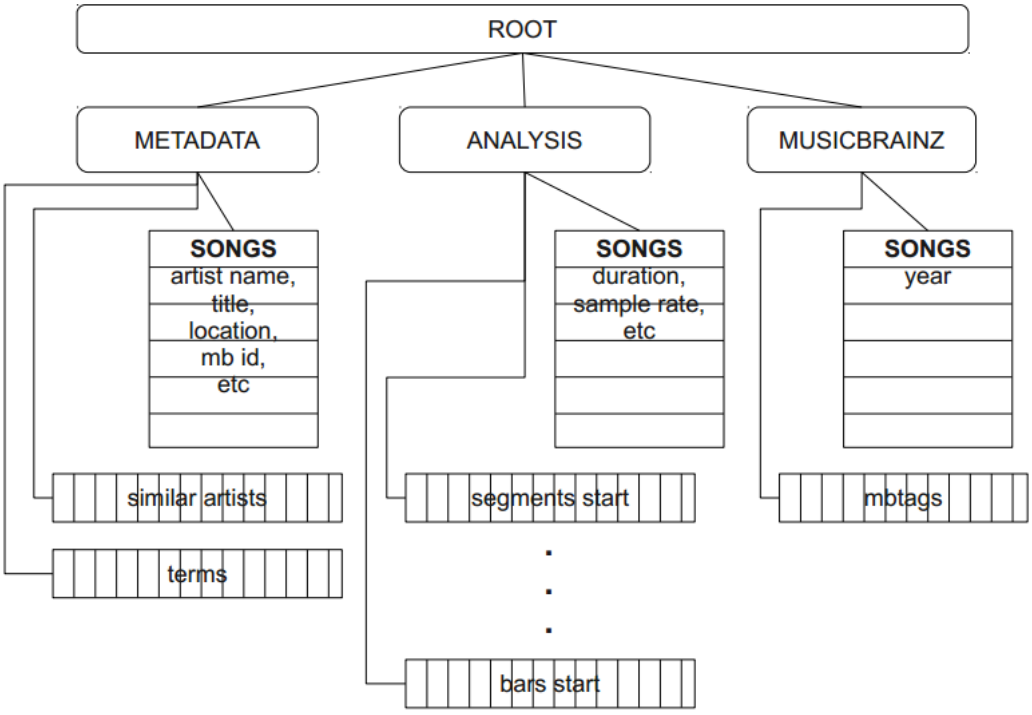

Figura A.1: Estructura de un archivo de MSD en formato HDF5.
Muestra la estructura del conjunto de tablas para los atributos de la canción.

A.3.2 Ritmo

Los *Beats* son subdivisiones de las barras. Los *Tatums* son subdivisiones de los *beats*. Es decir, las barras siempre se alinean con un *beat* y de la misma forma con los *Tatums*.

Se debe tener en cuenta que una confianza baja no significa necesariamente que el valor es incorrecto. Una confianza de -1 indica "no vale", por lo que el elemento correspondiente debe ser desechado. Una pista puede resultar sin *barra*, sin *beats*, y/o sin *Tatum* si no se detectó ninguna periodicidad. La signatura del tiempo (compás) oscila del 3 al 7 indicando compases de 3/4 a 7/4. Un valor de -1 puede indicar que no hay compás, mientras que un valor de 1 indica un compás bastante complejo o cambiante.

A.3.3 Clave

La clave (*Key*) es un atributo de nivel de la pista que va de 0 a 11 y que corresponde a una de las 12 notas: C, C#, D, etc. hasta B. Si no se ha detectado ninguna clave, el valor es -1.

A.3.4 Modo

El modo es igual a 0 o 1 para "menor" o "mayor" y puede ser -1 en caso de ningún resultado. Tenga en cuenta que la clave principal (por ejemplo, C mayor) más probable podría ser confundida con el tono menor a los 3 semitonos más bajos (por ejemplo, un A menor) ya que ambas claves tienen los mismos tonos.

A.3.5 Segmentos

Los segmentos (*segments*) son un conjunto de entidades de sonido (normalmente menores a un segundo de tiempo) cada uno relativamente uniforme en timbre y

armonía. Los segmentos se caracterizan por sus inicios perceptuales y duración en segundos, de sonoridad (dB), tono y contenido tímbrico. Más allá de la información de tiempo (inicio, duración), los segmentos incluyen características de sonoridad, tono y timbre.

A.3.5.1 **Sonoridad**

La información de sonoridad (ataque, decaimiento) viene dada por tres puntos de datos, incluyendo el valor dB de inicio (*loudness start*), el valor dB pico (*loudness max*) y la compensación (*offset*) relativo al segmento por la sonoridad máxima (*loudness max time*). El valor dB de aparición es equivalente al valor de dB en el desplazamiento para el segmento anterior. El último segmento también especifica un valor dB en el offset (*loudness end*).

A.3.5.2 **Tono**

El contenido de tono está dado por un vector "croma", que corresponde a la clase de los 12 tonos C, C#, D a B, con valores que van de 0 a 1 que describen el dominio relativo de cada lanzamiento en la escala cromática. Por ejemplo un acorde de Do Mayor probablemente estaría representada por grandes valores de C, E y G (es decir, clases 0, 4 y 7). Los vectores se normalizan a 1 por su dimensión más fuerte, por lo tanto, los sonidos ruidosos son probablemente representados por valores que están todos cerca de 1, mientras que los tonos puros son descritos por un valor a 1 (el tono) y otros cerca de 0.

A.3.5.3 **Timbre**

Es la cualidad de una nota musical o sonido que distingue a diferentes tipos de instrumentos musicales o voces. Es una noción compleja también denominada como color del sonido, textura o calidad de tono, y se deriva de la forma de la superficie espectro-temporal de un segmento, independientemente del tono y el volumen.

La característica timbre del Analizer de Echo Nest es un vector que incluye 12 valores no acotados más o menos centrados en 0. Estos valores son abstracciones de alto nivel de la superficie espectral, ordenados por grado de importancia. Sin embargo por integridad, la primera dimensión representa la sonoridad promedio del segmento; la segunda ilustra el brillo; la tercera se correlaciona más estrechamente con que tan plano es un sonido; la cuarta para sonidos con un ataque más fuerte, etc.

En la figura A.2 se representa a las 12 funciones base (es decir, segmentos plantilla). El timbre real del segmento se describe mejor como una combinación lineal de estas 12 funciones base ponderada por los valores de los coeficientes: $timbre = c_1 b_1 + c_2 b_2 + ... + c_{12} b_{12}$, donde c_1 a c_{12} representan los 12 coeficientes y b_1 a b_{12} las 12 funciones base.

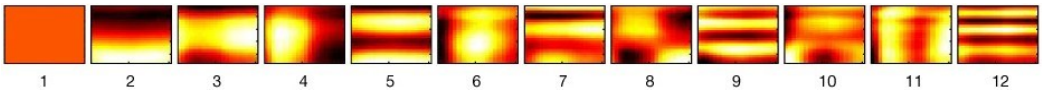

Figura A.2: Descripción de Segments_Timbre
12 funciones base para el vector timbre: x = tiempo, y = frecuencia, z = amplitud.

La siguiente lista complementa el conjunto de atributos analysis $\longrightarrow$ song de la tabla A.3. Representan los índices de cada uno de los atributos (tipo arreglo) de la tabla A.2. Son del tipo entero y tienen como valor por defecto 0.

Root⟶metadata⟶		
Atributo	Tipo	Descripción
artista_terms	Array string	Etiquetas de Echo Nest. Términos más descriptivos de un artista.
artista_terms_freq	Array float	Frecuencia de cada uno de los términos descriptivos de un artista (0.0 – 1.0).
artista_terms_weight	Array float	Pesos de cada uno de los términos descriptivos de un artista (0.0 – 1.0).
similar_artist	Array string	IDs de artistas similares en Echo Nest.
Root⟶metadata⟶songs⟶		
Atributo	Tipo	Descripción
analiyzer_version		
artist_7digitalid	int	ID de 7digital.com o -1 si no existe.
artist_familiarity	float	Estimación numérica de lo familiar que es un artista actualmente (0.0 – 1.0).
artist_hotttnesss	float	Estimación numérica de que tan hottt es un artista actualmente (0.0 – 1.0).
artist_id	string	ID de Echo Nest
artist_latitude	float	Latitud
artist_location	string	Nombre del lugar
artist_longitude	float	Longitud
artist_mbid	string	ID de MusicBrainz
artist_name	string	Nombre del artista
artist_playmeid	int	ID de playme.com o -1 si no existe.
genre	string	Etiqueta de género musical
idx_artist_terms	int	Índice del atributo artist_terms
idx_similar_artist	int	Índice del atributo similar_artist
release	string	Nombre del álbum de donde se tomó la pista, algunas canciones / la pistas puede venir de muchos álbumes, dan sólo uno.
release_7digitalid	int	ID de la liberación (álbum) en el servicio 7digital.com o -1 si no existe.
song_hotttnesss	float	Estimación numérica de que tan hottt es un canción actualmente (0.0 – 1.0).
song_id	string	ID de la canción en Echo Nest
title	string	Título de la canción
track_7digitalid	int	ID de 7digital.com o -1 si no existe

Figura A.3: Descripción de los atributos de MSD (Metadata)

Root⟶analysis⟶		
Atributo	Tipo	Descripción
bars_confidence	Array float	Medida de confianza para cada una de las barras localizadas.
bars_start	Array float	Indica el inicio de cada una de las barras, por lo general coincide con un beat. Sus valores representan el momento en que fueron localizadas (en segundos).
beats_confidence	Array float	Medida de confianza para cada uno de los beats localizados.
beats_start	Array float	Indica el inicio de cada una de los beats. Sus valores representan el momento en que se detectaron sobre la canción (en segundos).
sections_confidence	Array float	Medida de confianza para cada una de las secciones localizadas.
sections_start	Array float	Indica el inicio de cada una de las secciones (Agrupación más grande de una canción, por ejemplo: verso).Sus valores representan el momento en que se detectaron sobre la canción (en segundos).
segments_confidence	Array float	Medida de confianza para cada uno de los segmentos.
segments_loudness_max	Array float	Valor de sonoridad pico dentro de cada segmento (dB).
segments_loudness_max_time	Array float	Momento de la sonoridad máxima durante cada segmento.
segments_loudness_start	Array float	Indica el nivel de sonoridad en el inicio del segmento (dB).
segments_pitches	Array 2D float	Características croma un valor normalizado por cada nota. Para cada uno de los segmentos. (0.0 - 1.0)
segments_start	Array float	Eventos musicales, relacionados con los inicios de notas.
segments_timbre	Array 2D float	Características de textura (MFCC + PCA-like) para cada segmento.
tatums_confidence	float	Medida de confianza para cada uno de los tatums.
tatums_start	Array float	Lista de marcadores Tatum (en segundos). Tatums representan el tren de pulsos regulares más bajo que un oyente infiere intuitivamente desde el tiempo de los eventos musicales percibidos (segmentos).

Figura A.4: Descripción de los atributos de MSD (Analysis)

Root⟶analysis⟶song⟶		
Atributo	Tipo	Descripción
analysis_sample_rate	float	Frecuencia de muestreo del audio utilizado.
audio_md5	string	Código hash del audio utilizado por el análisis por Echo Nest. Número hexadecimal de 32 caracteres.
danceability	float	Medida de indica que tan bailable es una canción según Echo Nest (0.0 – 1.0, 0 sino es analizada)
duration	float	Duración de una pista calculada por el decodificador de audio (en segundos).
end_of_fade_in	float	Final del fundido de entrada: Indica el final de la introducción del fundido de entrada de la pista (en segundos).
energy	float	Medida de energía (no en el sentido de procesamiento de señales, desde el punto de vista del oyente de acuerdo con Echo Nest (0.0 – 1.0, 0 sino es analizada)
key	int	Es la clave global estimada de la pista. La clave identifica la tríada tónica, la cuerda, mayor o menor, que representa de forma definida el resto de la pieza. 0(Do) – 11(Si).
key_confidence	float	Medida de confianza para key.
loudness	float	Sonoridad: la sonoridad general de una pista en decibelios (dB). Los valores de volumen se promedian en el analizador cruzando la pista entera y son útiles para comparar la intensidad relativa de los segmentos y pistas. Loudness es la calidad de un sonido que es la correlación psicológica primaria de la fuerza física (amplitud).
mode	int	Indica la modalidad (mayor o menor) de una pista, el tipo de escala de la cual se deriva su contenido melódico.
mode_confidence	float	Medida de confianza para mode.
start_of_fade_out	float	Inicio de fundido de salida: el comienzo del fundido de salida al final de una pista en cuestión de segundos.
tempo	float	Tempo: el tempo global estimado de una pista en beats por minuto (BPM). En la terminología musical, el tempo es la velocidad o el ritmo de una pieza dada y se deriva directamente de la duración media.
time_signature	int	Estimado del compás sobre toda la pista. Es una notación que indica el número de beats por barra (3-7, 1 o -1).
time_signature_confidence	float	Medida de confianza para el compás (time signature).
track_id	string	ID de la pista en Echo Nest

Figura A.5: Descripción de los atributos de MSD (Analysis/Song)

Root⟶musicbrainz⟶		
Atributo	Tipo	Descripción
artist_mbtags	Array string	Etiquetas relacionadas con el artista en musicbrainz.
artist_mbtags_count	Array int	Cuenta las etiquetas de MusicBrainz
Root⟶musicbrainz⟶song⟶		
Atributo	Tipo	Descripción
idx_artist_mbtags	int	0 por defecto.
Year	int	Año de lanzamiento de la canción de MusicBrainz o 0 si no existe.

Figura A.6: Descripción de los atributos de MSD (MusicBrainz)

idx_bars_confidence
idx_bars_start
idx_beats_confidence
idx_beats_start
idx_sections_confidence
idx_sections_start
idx_segments_confidence
idx_segments_loudness_max
idx_segments_loudness_max_time
idx_segments_loudness_start
idx_segments_pitches
idx_segments_start
idx_segments_timbre
idx_tatums_confidence
idx_tantums_start

Figura A.7: La siguiente lista complementa el conjunto de atributos analysis ⟶ song de la figura A.5. Representan los índices de cada uno de los atributos (tipo arreglo) de la figura A.4. Son del tipo entero y tienen como valor por defecto 0 (MusicBrainz).

APÉNDICE

B

Redes de Creencia Profunda (DBN)

Las Redes de Creencia Profunda o DBN por sus siglas en inglés son un tipo de red que hasta hace algunos años no se utilizaban debido a la dificultad para entrenarlas. Estas son utilizadas en lo que se conoce como arquitecturas profundas. Algunos autores sugieren que son mucho más eficientes, en términos de los elementos computacionales requeridos, que las arquitecturas simples o de poca profundidad, como las redes neuronales con una sola capa oculta y las máquinas de soporte vectorial entre otras [23]. Sin embargo el entrenamiento de este tipo de redes es muy difícil, para superar este problema en el 2006 se introdujo un algoritmo de aprendizaje no supervisado que facilitó el entrenamiento [15].

En [24] se mostró que las maquinas restringidas de Boltzmann (RBM por sus siglas en inglés) se pueden apilar y entrenar de una manera voraz para formar DBNs. Las DBNs son modelos gráficos que aprenden a extraer representaciones jerárquicas profundas de los datos de entrenamiento. Modelan la distribución conjunta entre los valores observados el vector x y las ℓ capas ocultas, h^k, de la siguiente manera:

$$P(x, h^1, ..., h^\ell) = \left(\prod_{k=0}^{\ell-2} P(h^k | h^{k+1}) \right) P(h^{\ell-1}, h^\ell)$$

donde $x = h^0$, $P(h^{k-1}|h^k)$ es una distribución condicional de las unidades visibles condicionada a las unidades ocultas de la RBM en el nivel k, y $P(h^{\ell-1}, h^\ell)$ es la distribución conjunta visible oculta en el nivel superior de la RBM. Esto se ilustra en la figura siguiente.

Los parámetros de una DBN son los pesos W_j entre las unidades de las capas $j-1$ y j y el sesgo b_j de la capa j.

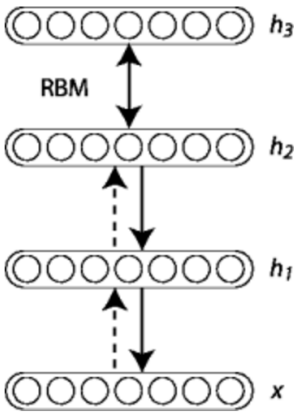

Figura B.1: Muestra el flujo de la información a traves de las capas

El principio de entrenamiento no supervisado de forma voraz de la capa oculta se puede aplicar a DBNs con RBMs como bloques de construcción para cada capa [23], [24]. El proceso es el siguiente:

1. Entrenar la primera capa como una RBM que modela la entrada bruta $x = h^{(0)}$ como su capa visible.

2. Usar esa primer capa para obtener una representación de la entrada que se utilizara como datos para la segunda capa. Existen dos soluciones comunes. Esta representación puede ser elegido como las activaciones promedio$p(h^{(1)} = 1|h^{(0)})$ o muestras de $p(h^{(1)}|h^{(0)})$.

3. Entrenar la segunda capa como RBM, tomando los datos transformados (ejemplos o activaciones promedio) como ejemplos de entrenamiento para la capa visible de esa RBM.

4. Iterar (2 y 3) para el número deseado de capas, haciendo la propagación hacia arriba.

5. finalmente ajustar con precisión todos los parámetros de esta arquitectura profunda con respecto a un indicador de la probabilidad log-DBN, o con respecto a un criterio de entrenamiento supervisado (después de la adición de las máquinas de aprendizaje adicionar por ejemplo, un clasificador lineal para convertir la representación aprendida en las predicciones supervisadas).

APÉNDICE

C

Maquinas de Boltzmann Restringidas (RBM)

La máquina de Boltzmann (BM por sus siglas en ingles), fue desarrollada por Geoffrey Hinton y Terry Sejnowski en 1983 [25]. Es un tipo de red neuronal donde todas las neuronas están conectadas entre si y tiene la particularidad de que toma decisiones estocásticas sobre si una neurona estará activada o no, es decir, son construidas introduciendo variaciones probabilistas a los pesos de la red. A esta máquina se le presenta un conjunto de vectores de entrenamiento que deberá aprender a clasificar con alta probabilidad, para lograrlo la BM debe encontrar los pesos de las conexiones que logren que los vectores de datos con los que fue entrenada presenten un "costo" o valor bajo en relación con otros ejemplos.

C.1 Modelos Basados en Energía (MBE).

Los modelos de energía asocian un escalar a la energía para cada configuración de las variables de interés. El Aprendizaje corresponde en modificar la función de energía de manera que su forma tenga propiedades deseables. Por ejemplo, nos

gustaría que las configuraciones posibles o deseables tengan baja energía. Los modelos probabilísticos basados en la energía definen una distribución de probabilidad a través de una función de energía, de la siguiente manera:

$$p(x) = \frac{e^{-Kx}}{Z}. \tag{C.1}$$

El factor de normalización Z es conocida como **función partición** por analogía con los sistemas físicos.

$$Z = \sum_{x} e^{-Ex}$$

Un modelo basado en energía puede aprenderse utilizando gradiente descendiente se puede aprender mediante la realización (estocástico) de descenso de gradiente sobre el negativo del logaritmo de la verosimilitud de los datos de entrenamiento. En cuanto a la regresión logística vamos primero a definir la verosimilitud y entonces la función de pérdida como el negativo de la verosimilitud.

$$\mathcal{L}(\theta, \mathcal{D}) = \frac{1}{N} \sum_{x^{(i)} \in \mathcal{D}} \log p(x^{(i)})$$
$$\ell(\theta, \mathcal{D}) = -\mathcal{L}(\theta, \mathcal{D})$$

usando el gradiente estocástico

$$-\frac{\partial \log p(x^{(x)})}{\partial \theta}$$

donde θ son los parámetros de el modelo.

C.1.1 MBE con Unidades Ocultas

En muchos casos de interés, no observamos el ejemplo x totalmente o queremos introducir algunas variables no observadas para mejorar el poder de expresión del modelo. Por lo que consideramos una parte observada (aún denotado x aquí) y una parte oculta h. Entonces podemos escribir:

$$P(x) = \sum_{h} P(x, h) = \sum_{h} \frac{e^{-E(x,h)}}{Z} \tag{C.2}$$

En tales casos, para mapear esta formula a una similar a la ecuación C.1, inspirados en la física se introduce la notación de la energía libre, definida de la siguiente manera:

$$\mathcal{F}(x) = -\log \sum_{h} e^{-E(e,h)} \tag{C.3}$$

lo cual nos permite reescribir

$$P(x) = \frac{e^{-\mathcal{F}(x)}}{Z} \textbf{ con } Z = \sum_{x} e^{-\mathcal{F}(x)}.$$

Los datos del gradiente del logaritmo negativo de la verosimilitud tienen una forma particularmente interesante.

$$-\frac{\partial \log p(x)}{\partial \theta} = \frac{\partial \mathcal{F}(x)}{\partial \theta} - \sum_{\tilde{x}} p(\tilde{x}) \frac{\partial \mathcal{F}(\tilde{x})}{\partial \theta}. \tag{C.4}$$

Es de notar que el gradiente de arriba contiene dos términos, los cuales son referenciados como la fase positiva y negativa. El término positivo y negativo no se refiere al signo de cada término en la ecuación, más bien reflejan el efecto sobre la densidad de probabilidad defina por el modelo. El primer término incrementa la probabilidad de los datos de entrenamiento, por la reducción de la energía correspondiente, mientras el segundo término decrece la probabilidad de los ejemplos generados por el modelo.

En realidad es difícil determinar el gradiente analíticamente, ya que involucra el cálculo de $E_P[\frac{\partial \mathcal{F}(x)}{\partial \theta}]$. Esto no es más que la esperanza sobre todas las configuraciones posibles de la entrada x bajo la distribución P formada por el modelo.

El primer paso para hacer este cálculo tratable es estimar la expectativa usando un número fijo de muestras modelo. Las muestras utilizadas para estimar el gradiente de fase negativa se denominan partículas negativos, que se denotan como $\mathcal{N}$. El gradiente puede entonces escribirse como:

$$\frac{\partial \log p(x)}{\partial \theta} \approx \frac{\partial \mathcal{F}(x)}{\partial \theta} - \frac{1}{|\mathcal{N}|} \sum_{\tilde{x} \in \mathcal{N}} \frac{\partial \mathcal{F}(\tilde{x})}{\partial \theta}. \tag{C.5}$$

donde nos gustaría idealmente elementos $\tilde{x}$ de $\mathcal{N}$ para ser muestras de acuerdo a P (es decir que estamos haciendo Montecarlo). Con la fórmula anterior, casi tenemos un algoritmo estocástico práctico, para el aprendizaje de un MBE. El único ingrediente que falta es cómo extraer estas partículas negativas $\mathcal{N}$. Si bien la literatura estadística está lleno de métodos de muestreo, los métodos de la cadena de Markov Monte Carlo son especialmente adecuados para los modelos tales como las RBM, un tipo específico de MBE.

C.2 Máquinas de Boltzmann Restringidas (RBM).

Las Máquinas de Boltzmann (BMs) son una forma particular de Campo Aleatorio de Markov (CAM) log-lineal, donde la función de energía es lineal en sus parámetros libres. Para que sean lo suficientemente potente como para representar distribuciones complicadas, es decir, pasar de la configuración paramétrica limitada a una no paramétrica, consideramos que algunas de las variables nunca son observadas (llamadas ocultas). Al tener más variables ocultas (también llamadas unidades ocultas), podemos aumentar la capacidad de modelado de la máquina de Boltzmann. Las RBM son aquellas BMs que no tienen conexiones de unidades visible-visible y oculta-oculta. La representación gráfica de una RBM se muestra en la figura C.1.

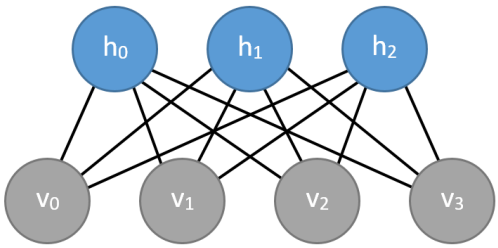

Figura C.1: Maquina de Boltzmann Restringida .
Existen sólo conexiones entre las unidades h_i y v_i.

La función de energía $E(v.h)$ de una MBR esta definida como:

$$E(v, h) = -b'vc' - h'Wv \tag{C.6}$$

donde W representa los pesos de las conexiones entre las unidades ocultas y visibles y b, c son las compensaciones de las capas visible y oculta respectivamente.

Esto se traduce directamente a la siguiente fórmula de la energía libre:

$$\mathcal{F}(v) = -b'v - \sum_i log \sum_{h_i} e^{h_i(c_i + W_i v)}$$

Dado que la estructura específica de la RBM, unidades visibles y ocultas son condicionalmente independientes una de otra. Utilizando esta propiedad, podemos escribir:

$$p(h|v) = \prod_i p(h_i|v) p(v|h) = \prod_j p(v_j|h)$$

C.2.1 RBMs con Unidades Binarias

En los casos comúnmente estudiados sobre el uso de unidades binarias, donde v_j y $h_i \in 0, 1$, obtenemos de las ecuaciones C.2y C.6, una versión probabilista de la función de activación de la neurona habitual:

$$P(h_i = 1|v) = sigm(c_i + W_i v) \quad \text{(C.7)}$$

$$P(v_j = 1|h) = sigm(b_j + W'_j h) \quad \text{(C.8)}$$

La energía libre de una RBM con unidades binarias se simplifica aún más:

$$\mathcal{F}v = -b'v - \sum_i log(1 + e^{(c_i + W_i v)}) \quad \text{(C.9)}$$

C.2.2 Ecuaciones de Actualización con Unidades Binarias

Combinando la ecuación C.5 con C.9, se obtienen los siguientes gradientes del logaritmo de verosimilitud para una RBM con unidades binarias.

$$
\begin{aligned}
-\frac{\partial \log p(v)}{\partial W_{ij}} &= E_v[p(h_i|v) \cdot v_j] - v_j^{(i)} \cdot sigm(W_i \cdot v^{(i)} + c_i) \\
-\frac{\partial \log p(v)}{\partial c_i} &= E_v[p(h_i|v)] - sigm(W_i \cdot v^{(i)}) \\
-\frac{\partial \log p(v)}{\partial b_j} &= E_v[p(v_j|h)] - v_j^{(i)})
\end{aligned}
\tag{C.10}
$$

C.3 Muestreando en una RBM.

Las muestras de $p(x)$ se pueden obtener ejecutando una cadena de Markov hasta su convergencia, mediante el muestreo de Gibss como el operador de transición.
El muestreo de Gibbs de la unión de N variables aleatorias $S = (S_1, ..., S_N)$ se realiza a través de una sucesión de N sub etapas de muestreo de la forma $S_i \sim p(S_i|S_{-i})$, donde S_{-i} contiene las otras $N-1$ variables aleatorias en S excluyendo S_i.
Para las RBMs, S consiste en el conjunto de las unidades visibles y ocultas. Sin embargo, ya que son condicionalmente independientes, uno puede realizar un bloque de muestreo de Gibbs. En esta configuración, las unidades visibles se muestrean simultáneamente dados los valores fijos de las unidades ocultas. Del mismo modo, las unidades ocultas se muestrean simultáneamente dadas las visibles. Un paso en la cadena de Markov se toma como sigue:

$$
\begin{aligned}
h^{(n+1)} &\sim sigm(W' v^{(n)} + c) \\
v^{(n+1)} &\sim sigm(W h^{(n+1)} + b)
\end{aligned}
$$

donde $h^{(n)}$ hace referencia a el conjunto de todas las unidades ocultas en el n paso de la cadena de Markov. Esto significa que, por ejemplo, $h_i^{(n+1)}$ es elegido de forma aleatoria para ser 1 (versus 0) con probabilidad $sigm(W_i' v^{(n)} + c_i)$ y de forma similar $v_j^{(n+1)}$ es elegida de forma aleatoria para ser 1 (versus 0) con probabilidad $sigm(W_j h^{(n+1)} + b_j)$.

Esto puede ilustrarse gráficamente:
Mientras $t \to \infty$, las muestras $((v(t), h(t))$ están garantizadas que son muestras precisas de $p(v, h)$.

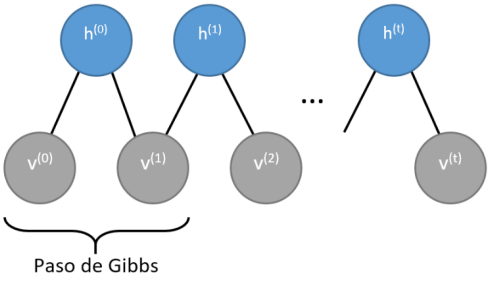

Figura C.2: Paso de Gibbs.

En teoría cada actualización de parámetros dentro del proceso de aprendizaje necesitaría ejecutar tal cadena para la convergencia.

Hacer esto sería prohibitivamente caro. Como tal, varios algoritmos se han ideado para las RBMs con el fin de muestrear eficientemente a partir de $p(v, h)$ durante el proceso de aprendizaje.

C.3.1 Divergencia Contrastiva (DC-k)

La Divergencia Contrastiva usa dos trucos para acelerar el proceso de muestreo.

- como se quiere que $p(v) \approx p_{entrenamiento}(v)$ (la verdadera distribución subyacente de los datos), se inicia la cadena de Markov con un ejemplo de entrenamiento, es decir, a partir de una distribución que se espera sea cercana a p, por lo que la cadena va a estar cercana a la distribución final p al haber convergido.

- DC no espera a que la cadena converja. Las muestras se obtienen a partir de solamente k-etapas de muestreo de Gibbs. Se ha demostrado que en la práctica, $k = 1$ funciona muy bien.

C.3.2 DC Persistente

DC persistente usa otra aproximación para el muestreo a partir de $p(v, h)$. Se basa en una única cadena de Markov, que tiene un estado persistente (es decir, no reiniciar una cadena para cada ejemplo observado). Para cada actualización de paráme-

tros, extraemos nuevas muestras simplemente ejecutando la cadena de k-pasos. El estado de la cadena se conserva luego para actualización posterior.
La intuición general es que si las actualizaciones de los parámetros son suficientemente pequeñas en comparación con la taza de mezcla de la cadena, la cadena de Markov debe ser capaz de "ponerse al día" a los cambios en el modelo.

APÉNDICE

D

Divergencia Contrastiva

Divergencia Contrastiva (DC): Algoritmo de aprendizaje propuesto por Geoffrey Hinton para aproximar la Máxima Verosimilitud (MV)[1].

¿Qué es la Divergencia Contrastiva y para qué nos sirve?

Imagíne que nos gustaría modelar la probabilidad de un dato en un punto x utilizando una función de la forma $f(x;\Theta)$, donde Θ es un vector de parámetros del modelo. La probabilidad de x, $p(x;\Theta)$ debe ser 1 al integrar sobre toda x. Por lo tanto:

$$p(x:\Theta) = \frac{1}{Z(\Theta)} f(x;\Theta) \tag{D.1}$$

donde $Z(\Theta)$ es conocida como la función partición, y se define como:

$$Z(\Theta) = \int f(x;\Theta)dx \tag{D.2}$$

[1]Notes on Contrastive Divergence by Oliver Woodfort.
url: http://www.robots.ox.ac.uk/ ojw/files/NotesOnCD.pdf

Aprenderemos nuestros parámetros del modelo Θ, maximizando la probabilidad del conjunto de datos de entrenamiento $X = x_{1,\dots,K}$, dada como:

$$p(X;\Theta) = \prod_{k=1}^{K} \frac{1}{Z(\Theta)} f(x_k;\Theta) \tag{D.3}$$

o equivalentemente, por minimización del logaritmo negativo de $p(X;\Theta)$ denotado por $E(X;\Theta)$ al cual llamaremos Energía:

$$E(X;\Theta) = \log Z(\Theta) - \frac{1}{K}\sum_{k=1}^{K} \log f(x_k;\Theta) \tag{D.4}$$

Primero vamos ha elegir nuestro modelo de función de probabilidad $f(x;\Theta)$ de forma que sea la fdp de una distribución normal $\mathcal{N}(x;\mu,\sigma)$, tal que $\Theta = \{\mu,\sigma\}$. La integral de la fdp es igual a 1 (un reultado estandar, aunque la pruba no es trivial), de modo que $\log Z(\Theta) = 0$. Diferenciando la ecuación (D.4) con respecto a μ se muestra que la μ óptima es la media de los datos de entrenamiento X, y un calculo similar con respecto a σ muestra que la σ óptima es la raíz cuadrada de la varianza de los datos de entrenamiento.

Algunas veces, como en este caso, existe un método que puede exactamente minimizar nuestra función de energía en particular. Si imaginamos nuestra función de energía en un espacio de parámetros dentro de un campo ondulante cuyo punto más bajo queremos hallar, podríamos decir que este caso es el equivalente a estar en un día claro y soleado en el campo, al ver el punto más bajo y caminar directamente a él.

Ahora vamos a elegír nuestro modelo de la función de probabilidad $f(x;\Theta)$, para que sea la suma de N distribuciones normales, tal que $\Theta = \{\mu_{1,\cdots,N}, \sigma_{1,\cdots,N}\}$ y

$$f(x;\Theta) = \sum_{i=1}^{N} \mathcal{N}(x;\mu_i,\sigma_i) \tag{D.5}$$

Esto es equivalente a una suma de expertos o un modelo de mezclas con pesos iguales sobre todos los expertos; tener pesos diferentes es una extensión trivial para el modelo. Nuevamente usando el hecho que la integral de una distribución normal

es 1, podemos ver de la ecuación (D.2) que $\log Z(\Theta) = \log N$. Sin embargo ahora diferenciando la ecuación (D.4) con respecto a cada uno de nuestros parámetros del modelo se producen ecuaciones que dependen de otros parámetros del modelo, por lo que no podemos calcular los parámetros óptimos del modelo directamente. En su lugar podemos utilizar ecuaciones diferenciales parciales y un método de gradiente descendente con búsqueda lineal para encontrar un mínimo local de energía en el espacio de parámetros.

Volviendo a nuestra metáfora del campo, podríamos decir que el gradiente descendente con la búsqueda lineal equivale a estar en el campo por la noche con una linterna. Podemos sentir el gradiente del campo en el punto donde estamos de pie, o bien mediante el uso de la linterna estimar la altura relativa viendo el campo a una corta distancia de nosotros en cada sentido (diferenciación numérica usando diferencias finitas). Entonces, el resplandor de la luz emitida por la linterna en la dirección elegida para el viaje, también nos permite ver el punto más bajo en el campo en esa dirección. A continuación, podemos caminar hasta ese punto, y de forma iterativa elegir una nueva dirección y la distancia a caminar.

Finalmente vamos a elegir como modelo nuestra función de probabilidad $f(x;\Theta)$, el producto de N distribuciones normales, tal que

$$f(x;\Theta) = \prod_{i=1}^{N} \mathcal{N}(x;\mu_i,\sigma_i) \tag{D.6}$$

Esto es equivalente al producto de modelos expertos. La función partición $Z(\Theta)$ ahora ya no es una constante. Podemos observar esto al considerar un modelo que consista de dos distribuciones normales, ambas con $\sigma = 1$. Si $\mu_1 = \infty$ y $\mu_2 = \infty$ entonces $Z(\Theta) = 0$, mientras que si $\mu_1 = \mu_2 = 0$ entonces $Z(\Theta) = 1/2\sqrt{\pi}$.

Si bien esto es posible, en este caso para calcular la función partición exactamente dada Θ, vamos a imaginar que la integral de la ecuación (D.2) no es algebraicamente tratable (como sería el caso con otro modelo de función de probabilidad). En este caso podríamos necesitar el uso de un método de intregación numérico para evaluar la ecuación (D.4), usar diferencias finitas para calcular el gradiente en

un punto dado en el espacio de parámetros y use un método de gradiente descendiente para encontrar un mínimo local. Para espacios de datos de gran dimensión el tiempo de integración es paralizante, y un espacio de parámetros de gran dimensión provoca este problema. Esto nos lleva a una situación en la que estamos tratando de minimizar una función de energía que no podemos evaluar.

Aquí es donde la DC nos ayuda. A pesar de que no podemos evaluar la función de energía en sí, DC proporciona una manera de estimar el gradiente de la función de energía. Si regresamos a nuestra metáfora del campo, ahora nos encontramos en el campo sin ninguna luz en absoluto (es decir, no podemos calcular la energía), por lo que no se puede establecer la altura de cualquier punto en el campo con relación a la nuestra. DC nos da efectivamente un sentido de equilibrio, lo que nos permite sentir el gradiente del campo bajo nuestros pies. Al tomar pasos muy pequeños en la dirección del gradiente más inclinado podemos hallar nuestro camino a un mínimo local.

¿Cómo trabaja DC?

Como se explicó, DC estima nuestra función gradiente de energía, dando un conjunto de parámetros del modelo Θ y nuestros datos de entrenamiento X. Podríamos derivar la ecuación del gradiente primeramente escribiendo la derivada parcial de la ecuación (D.4):

$$\frac{\partial E(X;\Theta)}{\partial \Theta} = \frac{\partial \log Z(\Theta)}{\partial \Theta} - \frac{1}{K}\sum_{i=1}^{K}\frac{\partial \log f(x_i;\Theta)}{\partial \Theta} \tag{D.7}$$

$$= \frac{\partial \log Z(\Theta)}{\partial \Theta} - \left\langle \sum_{i=1}^{K}\frac{\partial \log f(x_i;\Theta)}{\partial \Theta} \right\rangle_X \tag{D.8}$$

donde $\langle\cdot\rangle_X$ es la esperanza de la distribución de los datos dados X.

El primer término de la derecha viene de la función partición, la cual como muestra la ecuación (D.2), involucra una integración sobre x. Substituyéndo en esta, obtenemos:

$$\frac{\partial \log Z(\Theta)}{\partial \Theta} = \frac{1}{Z(\Theta)} \frac{\partial Z(\Theta)}{\partial \Theta} \tag{D.9}$$

$$= \frac{1}{Z(\Theta)} \frac{\partial}{\partial \Theta} \int f(x;\Theta) dx \tag{D.10}$$

$$= \frac{1}{Z(\Theta)} \int \frac{\partial f(x;\Theta)}{\partial \Theta} dx \tag{D.11}$$

$$= \frac{1}{Z(\Theta)} \int f(x;\Theta) \frac{\partial \log f(x_i;\Theta)}{\partial \Theta} dx \tag{D.12}$$

$$= \int p(x;\Theta) \frac{\partial \log f(x;\Theta)}{\partial \Theta} dx \tag{D.13}$$

$$= \left\langle \frac{\partial \log f(x;\Theta)}{\partial \Theta} \right\rangle_{p(x;\Theta)} \tag{D.14}$$

Como se ha expuesto, esta integración generalmente es intratable algebraicamente. Sin embargo, en la forma de la ecuación (D.14), es evidente que se puede aproximar numéricamente tomando muestras a partir de la distribución propuesta, $p(x;\Theta)$.

Las muestras no se pueden extraer directamente de $p(x;\Theta)$, ya que no conocemos el valor de la función de partición, pero podemos utilizar muchos ciclos de Cadenas de Markov Monte Carlo (MCMC) tomando muestras para transformar los datos de entrenamiento (extraídos de la distribución de destino) en datos extraídos de la distribución propuesta.

Esto es posible ya que la transformación sólo implica el cálculo de la relación de dos probabilidades, $p(x';\Theta)/p(x;\Theta)$, por lo que la función de partición se anula. X^n representa los datos de entrenamiento transformadas usando n ciclos de MCMC, tal que $X^0 \equiv X$. Sustituyendo en la ecuación (D.8), se obtiene:

$$\frac{\partial E(X;\Theta)}{\partial \Theta} = \left\langle \frac{\partial \log f(x;\Theta)}{\partial \Theta} \right\rangle_{X^\infty} - \left\langle \frac{\partial \log f(x;\Theta)}{\partial \Theta} \right\rangle_{X^0} \tag{D.15}$$

Todavía tenemos un obstáculo computacional por superar, el calculo de un gradiente preciso requiere de tantos ciclos-MCMC que se llevaría demasiado tiempo. La afirmación de Hinton fue que serían necesarios sólo unos pocos ciclos MCMC para calcular un gradiente aproximado. La intuición detrás de esto es que después

de unas pocas iteraciones los datos se han de moverse de la distribución destino (es decir, los datos de entrenamiento) hacia la distribución propuesta, y así dar una idea de en qué dirección la distribución propuesta debe moverse para un mejor modelo de los datos de entrenamiento. Empíricamente, Hinton ha encontrado que incluso 1 ciclo de MCMC es suficiente para que el algoritmo converja a la respuesta de ML (Máxima verosimilitud). Por lo tanto, teniendo en cuenta que deseamos ir cuesta abajo con el fin de reducir al mínimo nuestra función de energía, nuestra ecuación de actualización de parámetros se puede escribir como:

$$\Theta_{t+1} = \Theta_t + \eta \left(\left\langle \frac{\partial \log f(x;\Theta)}{\partial \Theta} \right\rangle_{X^\infty} - \left\langle \frac{\partial \log f(x;\Theta)}{\partial \Theta} \right\rangle_{X^0} \right) \tag{D.16}$$

donde η es el factor tamaño del paso, el cual debe ser elegido experimentalmente, basado en el tiempo de convergencia y estabilidad.

APÉNDICE

E

Timbre

El Timbre es uno de los factores más importantes del sonido, conocido también como el color del tono. El timbre representa las cualidades auditivas que hacen que un instrumento suene distinto a otro. Gracias a él, podamos percibir la diferencia entre un violín que resuena en la madera y tiene un toque cálido, y el sonido distorsionado de una guitarra eléctrica que normalmente es vivo y estridente (Figura E.1).

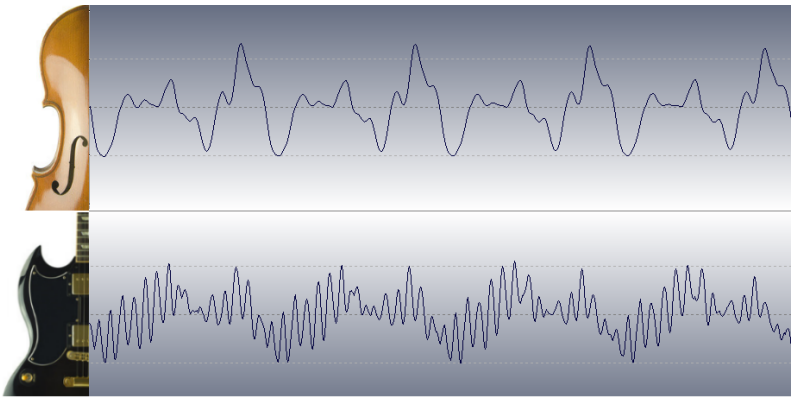

Figura E.1: Señales de violín y guitarra eléctrica.

El concepto de timbre no solo se limita a los instrumentos musicales, también es la razón por la cual podemos distinguir con quién se está hablando por teléfono o

identificar a un cantante conocido con solo escuchar una nota a capella.

El timbre permite distinguir a quien pertenece la misma nota emitida por distintos instrumentos.

Todos estos timbres distintos son posibles mediante la combinación de dos factores: las vibraciones del instrumento en sí mismo y las frecuencias que estas vibraciones producen. Tanto el tamaño como la forma del instrumento tienen un gran impacto en las vibraciones producidas.

La sinestesia, es un fenómeno mediante el cual las personas son capaces de ver colores cuando oyen determinados sonidos; el timbre es un importante factor a la hora de determinar cuáles son esos colores. Un instrumento de viento de sonido vivo, por ejemplo, podría describirse como amarillo, mientras que un timbal representaría un tono mucho más oscuro. Quizás es esto a lo que se refería John Lennon cuando pidió al productor George Martin que hiciera que una de las canciones en las que estaba trabajando "sonara como una naranja"[1].

Figura E.2: El color de la Música.
El timbre es quien nos hace pensar en colores o texturas.

[1]La importancia del timbre. https://community.sony.es/t5/blog-noticias-sony/la-importancia-del-timbre/ba-p/1916559

E.1 Armónicos.

Son los componentes de un sonido que se definen como las frecuencias secundarias que acompañan a una frecuencia fundamental o generadora.

Los sonidos armónicos son producidos por la naturaleza de los cuerpos capaces de vibrar al recibir ondas sonoras que emiten un sonido fundamental al espacio.

El armónico de una onda es un componente sinusoidal de una señal. Su frecuencia es un múltiplo de la fundamental.

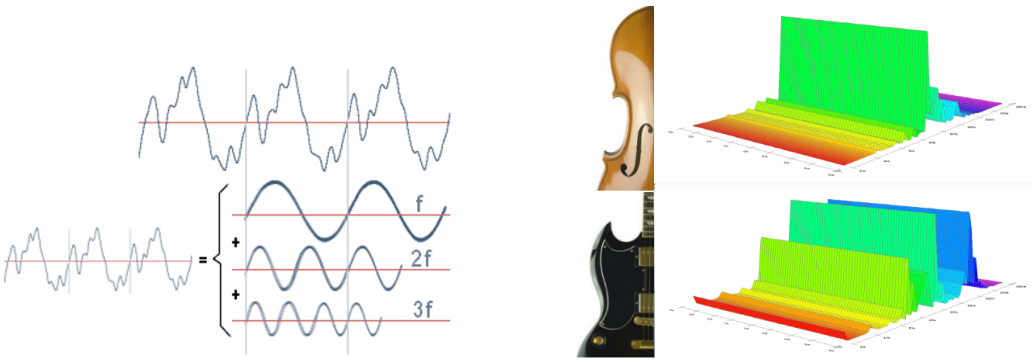

Figura E.3: La amplitud de los armónicos afecta el timbre.

En la figura E.3a se muestra la suma del 2^o y 3^{er} armónico sobre la fundamental. En la figura E.3b se observan dos ejemplos de como la diferencia en la amplitud de los armónicos describe el timbre de dos instrumentos ejecutando la nota de LA (440Hz).

La amplitud de los armónicos más altos es mucho menor que la amplitud de la onda fundamental y tiende a cero; por este motivo los armónicos por encima del quinto o sexto generalmente son inaudibles.

Los armónicos son a su vez, los que generan el timbre característico de una fuente de sonido. La amplitud y ubicación de los primeros armónicos es lo que nos permite distinguir diferentes timbres(instrumentos, voces, etc.).

Cuando se ejecuta una nota en un instrumento musical se genera una onda de presión de aire. Esta onda sonora está acompañada por una serie de armónicos, todos prácticamente inaudibles, pero que le dan al instrumento su timbre particular. Cada armónico de esta serie tiene una amplitud (volumen o fuerza del sonido) diferente. Por ejemplo en el clarinete son más fuertes los armónicos impares (el 3º, el 5º, el 7º, etc.).

A partir del quinto armónico, todos los siguientes armónicos impares suenan ligeramente desafinados con respecto al temperamento igual[1].

Los armónicos cuyas frecuencias no son múltiplos enteros se denominan "parciales". Las campanas son los que poseen más parciales perceptibles que otros instrumentos

E.2 Serie armónica.

Sucesión de los sonidos cuyas frecuencias son múltiplos enteros positivos de la nota base, llamada fundamental.

Para estudiar la serie armónica se numera cada sonido con un índice, comenzando por el número uno para el sonido fundamental. Una importante propiedad de la serie son las las razones o cocientes entre los índices respectivos de dos sonidos cualesquiera. Esta proporción caracteriza al mismo intervalo entre dos notas

[1]Temperamento igual: es el nombre común del sistema temperado de doce notas, que es el sistema de afinación más utilizado actualmente en la música occidental, y que se basa en el semitono temperado, igual a la doceava parte de la octava y de razón numérica igual a la raíz doceava de dos, con una amplitud de intervalo de 100 cents.

cualesquiera, cuando sus frecuencias se encuentran en la misma proporción (Ver la tabla E.1).

No. de Armónico	Frecuencia	Nota	Intervalo
1° armónico	264 Hz	do1	tono fundamental (el primer do a la izquierda del piano)
2° armónico	528 Hz	do2	octava
3° armónico	792 Hz	sol2	quinta
4° armónico	1056 Hz	do3	octava
5° armónico	1320 Hz	mi3	tercera mayor
6° armónico	1584 Hz	sol3	quinta
7° armónico	1848 Hz	sib3	séptima menor (muy desafinada)
8° armónico	2112 Hz	do4	octava
9° armónico	2376 Hz	re4	segunda mayor
10° armónico	2640 Hz	mi4	tercera mayor
11° armónico	2904 Hz	fa#4	cuarta aumentada
12° armónico	3168 Hz	sol4	quinta justa
13° armónico	3432 Hz	la4	sexta mayor (muy desafinada)
14° armónico	3696 Hz	sib4	séptima menor
15° armónico	3960 Hz	si4	séptima mayor
16° armónico	4224 Hz	do5	octava

Tabla E.1: Serie de los primeros armónicos principales.

E.2.1 El Papel de Cada Armónico

La contribución de cada armónico al timbre del sonido se puede ver como una receta donde cada armónico es un ingrediente.

EL sonido fundamental proporciona por sí solo la misma sensación de altura que el fundamental con todos sus armónicos; decimos que la frecuencia de la nota que se oye es igual a la del sonido fundamental.

El fenómeno de la “fundamental fantasma” tiene su explicación en el carácter no lineal del oído humano, el sonido fundamental no es imprescindible para percibir el conjunto como una nota con la misma altura, siempre y cuando existan o suenen el resto de los sonidos de la serie armónica. El oído reconstruye el sonido que falta como si dedujese este resultado de una ecuación cuya única solución posible es esta fundamental.

Los sonidos números 2, 4, 8 y todos los que forman una relación igual a una potencia de 2 con la fundamental, refuerzan el carácter inequívoco de la sensación de altura del conjunto. Los sonidos 3, 6, 12 y todos aquellos que forman con el 3 una relación que es una potencia de 2, aportan un timbre nasal al conjunto.
Los sonidos 5 y 10 producen un timbre o color “redondo”, “profundo”, “cálido” y otros adjetivos semejantes.

Los sonidos 7, 11, 13 y 15 son disonantes y dan un carácter “áspero” al sonido.

Al crecer el número de orden de un armónico, su aportación es de más brillantez o claridad; más brillantez que claridad si es un número múltiplo de los 16 primeros excepto los que hemos denominado como disonantes.

APÉNDICE

F

Experimentos K-means

Se utilizó K-means con las siguientes especificaciones:

- 4 agrupamientos o clusters,
- distancia euclidiana
- limite de 500 iteraciones
- se evaluó utilizando el género pre-asignado.

Se utilizaron las características tono o pitch (P), timbre (T) e intensidad (I).
Se utilizó un conjunto de configuraciones para observar cómo las características seleccionadas afectaban el agrupamiento.

$$Configuraciones = \{P, T, I, PT, PI, TI, PTI\}$$

En los experimentos del 1 al 4 se utiliza la Base de datos de la tabla F.1

F.1 Experimento 1.

Se utilizaron los cuatro primeros coeficientes del vector del timbre que están descritos en la documentación (Ver apéndice A).

Las configuraciones PTI, PT, PI y P agrupan casi la misma cantidad de ejemplos. Al reportarlos como porcentajes con un decimal, el redondeo hace que los resultados sean iguales (Tablas F.2, F.3, F.4 y F.5).

En la figura F.1 se muestra el porcentaje de ejemplos clasificados para el experimento 1.

Tabla F.1: Géneros y ejemplos de la BD-1.

Género Musical	No. de Ejemplos	% Ejemplos
Blues	170699	23 %
Folk	149231	20 %
Jazz	225089	30 %
Reggae	198292	27 %
Total	743311	100 %

Distribución de los ejemplos del conjunto de 200 canciones por géneros.

Tabla F.2: Experimento 1. Configuración PTI.

Género	C0	C1	C2	C3	%Contribución
Blues	6.1 %	6.4 %	5.4 %	5.0 %	23.0 %
Folk	5.6 %	6.7 %	4.8 %	3.0 %	20.1 %
Reggae	5.2 %	6.2 %	8.8 %	6.5 %	26.7 %
Jazz	7.2 %	7.8 %	7.9 %	7.4 %	30.3 %
% Ej. Agrupados	24.1 %	27.1 %	26.9 %	21.9 %	

Clasificación general: 29.0 %.

Tabla F.3: Experimento 1. Configuración PT.

Género	C0	C1	C2	C3	%Contribución
Blues	6.1 %	6.4 %	5.4 %	5.0 %	23.0 %
Folk	5.6 %	6.7 %	4.8 %	3.0 %	20.1 %
Reggae	5.2 %	6.2 %	8.8 %	6.5 %	26.7 %
Jazz	7.2 %	7.8 %	7.9 %	7.4 %	30.3 %
% Ej. Agrupados	24.1 %	27.1 %	26.9 %	21.9 %	

Clasificación general: 29.0 %.

Tabla F.4: Experimento 1. Configuración PI.

Género	C0	C1	C2	C3	%Contribución
Blues	6.1 %	6.4 %	5.4 %	5.0 %	23.0 %
Folk	5.6 %	6.7 %	4.8 %	3.0 %	20.1 %
Reggae	5.2 %	6.2 %	8.8 %	6.5 %	26.7 %
Jazz	7.2 %	7.8 %	7.9 %	7.4 %	30.3 %
% Ej. Agrupados	24.1 %	27.1 %	26.9 %	21.9 %	

Clasificación general: 29.0 %.

Tabla F.5: Experimento 1. Configuración P.

Género	C0	C1	C2	C3	%Contribución
Blues	6.1 %	6.4 %	5.4 %	5.0 %	23.0 %
Folk	5.6 %	6.7 %	4.8 %	3.0 %	20.1 %
Reggae	5.2 %	6.2 %	8.8 %	6.5 %	26.7 %
Jazz	7.2 %	7.8 %	7.9 %	7.4 %	30.3 %
% Ej. Agrupados	24.1 %	27.1 %	26.9 %	21.9 %	

Clasificación general: 29.0 %.

F. EXPERIMENTOS K-MEANS

Tabla F.6: Experimento 1. Configuración I.

Género	C0	C1	C2	C3	%Contribución
Blues	8.7 %	8.7 %	1.0 %	4.6 %	23.0 %
Reggae	10.3 %	11.0 %	1.0 %	4.4 %	26.7 %
Folk	5.9 %	7.6 %	1.4 %	5.2 %	20.1 %
Jazz	8.8 %	11.1 %	2.4 %	8.0 %	30.3 %
% Ej. Agrupados	33.7 %	38.3 %	5.8 %	22.2 %	

Clasificación general: 29.1 %.

Tabla F.7: Experimento 1. Configuración TI.

Género	C0	C1	C2	C3	%Contribución
Blues	10.9 %	3.7 %	1.9 %	6.4 %	23.0 %
Reggae	13.4 %	8.5 %	1.7 %	3.0 %	26.7 %
Jazz	11.1 %	6.9 %	4.8 %	7.5 %	30.3 %
Folk	7.1 %	4.3 %	3.0 %	5.7 %	20.1
% Ej. Agrupados	42.5 %	23.4 %	11.4 %	22.7 %	

Clasificación general: 29.9 %.

Tabla F.8: Experimento 1. Configuración T.

Género	C0	C1	C2	C3	%Contribución
Reggae	12.8 %	9.2 %	1.6 %	3.1 %	26.7 %
Jazz	9.7 %	7.4 %	5.2 %	7.9 %	30.3 %
Folk	5.8 %	4.7 %	3.4 %	6.2 %	20.1 %
Blues	9.8 %	3.9 %	2.1 %	7.1 %	23.0 %
% Ej. Agrupados	38.1 %	25.3 %	12.2 %	24.4 %	

Clasificación general: 30.7 %.

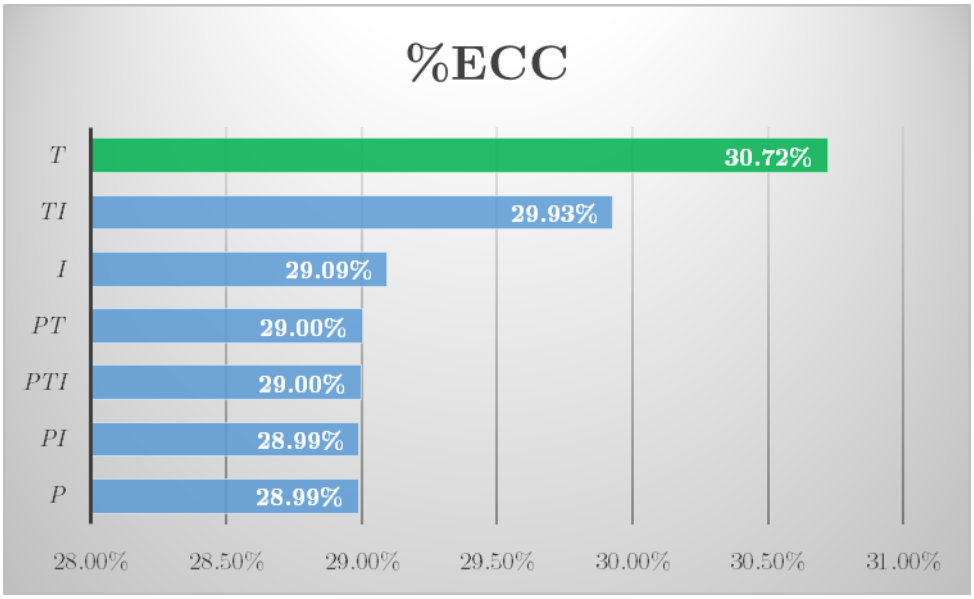

Figura F.1: %ECC:Porcentaje de ejemplos clasificados correctamente del experimento 1.

F.2 Experimento 2.

Se utilizó el total de coeficientes del vector del timbre.

Los resultados para las configuraciones que no involucran T son los mismos del experimento 1. Por lo tanto sólo se muestran las tablas de resultados para las configuraciones donde aparece T.

Las configuraciones PTI, PT, PI y P agrupan casi la misma cantidad de ejemplos. Al reportarlos como porcentajes con un decimal, el redondeo hace que los resultados sean iguales (Tablas F.4, F.5, F.9 y F.10).

En la figura F.2 se muestran el porcentaje de ejemplos clasificados para el experimento 2 y la comparación con el experimento 1.

Tabla F.9: Experimento 2. Configuración PTI.

Género	C0	C1	C2	C3	%Contribución
Blues	6.1 %	6.4 %	5.4 %	5.0 %	23.0 %
Folk	5.6 %	6.7 %	4.8 %	3.0 %	20.1 %
Reggae	5.2 %	6.2 %	8.8 %	6.5 %	26.7 %
Jazz	7.2 %	7.8 %	7.9 %	7.4 %	30.3 %
% Ej. Agrupados	24.1 %	27.1 %	26.9 %	21.9 %	

Clasificación general: 29.0 %.

Tabla F.10: Experimento 2. Configuración PT.

Género	C0	C1	C2	C3	%Contribución
Blues	6.1 %	6.4 %	5.4 %	5.0 %	23.0 %
Folk	5.6 %	6.7 %	4.8 %	3.0 %	20.1 %
Reggae	5.2 %	6.2 %	8.8 %	6.5 %	26.7 %
Jazz	7.2 %	7.8 %	7.9 %	7.4 %	30.3 %
% Ej. Agrupados	24.1 %	27.1 %	26.9 %	21.9 %	

Clasificación general: 29.0 %.

Tabla F.11: Experimento 2. Configuración TI.

Género	C0	C1	C2	C3	%Contribución
Reggae	14.0 %	8.0 %	1.6 %	3.1 %	26.7 %
Jazz	11.2 %	7.1 %	4.9 %	7.1 %	30.3 %
Folk	7.0 %	4.7 %	3.0 %	5.4 %	20.1 %
Blues	10.7 %	3.8 %	1.9 %	6.5 %	23.0 %
% Ej. Agrupados	42.8 %	23.7 %	11.4 %	22.1 %	

Clasificación general: 30.6 %.

Tabla F.12: Experimento 2. Configuración T.

Género	C0	C1	C2	C3	%Contribución
Reggae	13.5 %	8.7 %	1.5 %	2.9 %	26.7 %
Jazz	9.5 %	8.0 %	5.3 %	7.4 %	30.3 %
Folk	5.5 %	5.3 %	3.4 %	5.9 %	20.1 %
Blues	9.2 %	4.3 %	2.1 %	7.3 %	23.0 %
% Ej. Agrupados	37.7 %	26.3 %	12.4 %	23.6 %	

Clasificación general: 32.2 %.

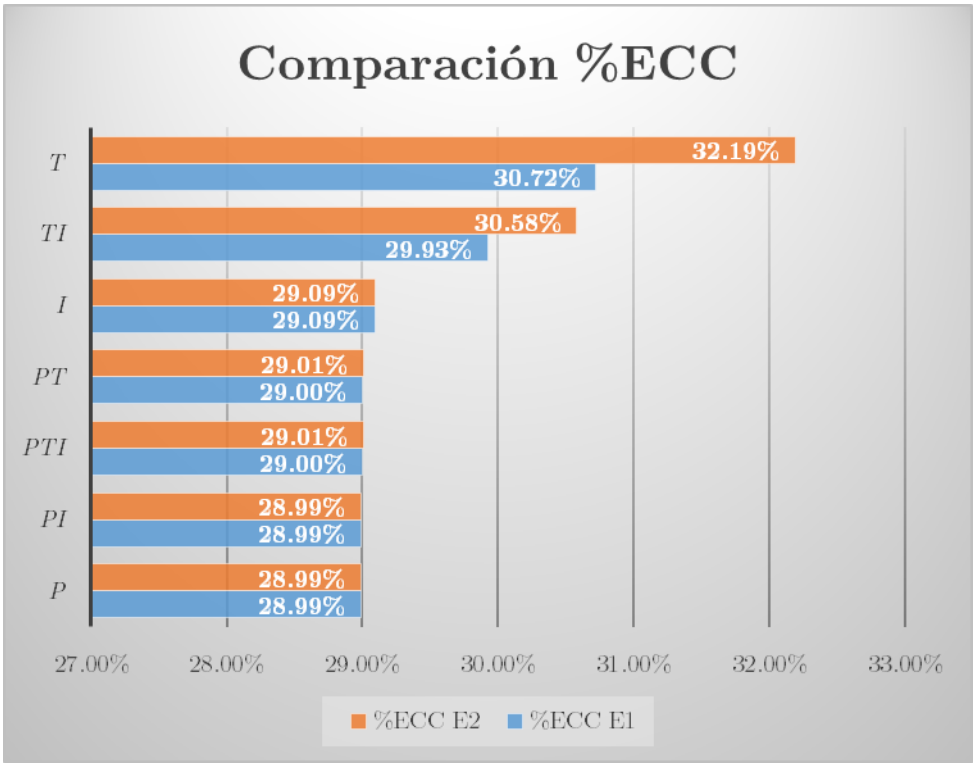

Figura F.2: Comparación de %ECC de los experimentos 1 y 2.

F.3 Experimento 3.

Se utilizaron los cuatro primeros coeficientes del vector del timbre y se balanceo la base igualando la cantidad de ejemplos por género, al número de ejemplos del género con menor cantidad, en este caso Folk (Tabla F.13).

Las configuraciones PTI, PT, PI y P agrupan casi la misma cantidad de ejemplos. Al reportarlos como porcentajes con un decimal, el redondeo hace que los resultados sean iguales (Tablas F.14, F.15, F.16 y F.17).

En la figura F.3 se muestran el porcentaje de ejemplos clasificados para el experimento 3 y la comparación con el experimento 1.

Tabla F.13: Géneros y ejemplos de la BD-1 balanceada.

Género Musical	No. de Ejemplos	% Ejemplos
Blues	149231	25 %
Folk	149231	25 %
Jazz	149231	25 %
Reggae	149231	25 %
Total	596924	100 %

Distribución de los ejemplos por géneros.

Tabla F.14: Experimento 3. Configuración PTI.

Género	C0	C1	C2	C3	%Contribución
Reggae	8.2 %	4.9 %	6.1 %	5.8 %	25.0 %
Blues	5.8 %	6.7 %	5.6 %	6.9 %	25.0 %
Jazz	6.4 %	5.8 %	6.1 %	6.7 %	25.0 %
Folk	5.9 %	7.0 %	3.8 %	8.3 %	25.0 %
% Ej. Agrupados	26.4 %	24.4 %	21.6 %	27.6 %	

Clasificación general: 29.3 %.

Tabla F.15: Experimento 3. Configuración PT.

Género	C0	C1	C2	C3	%Contribución
Reggae	8.2 %	4.9 %	6.1 %	5.8 %	25.0 %
Blues	5.9 %	6.7 %	5.6 %	6.9 %	25.0 %
Jazz	6.4 %	5.8 %	6.1 %	6.7 %	25.0 %
Folk	5.9 %	7.0 %	3.8 %	8.3 %	25.0 %
% Ej. Agrupados	26.4 %	24.4 %	21.6 %	27.6 %	

Clasificación general: 29.3 %.

Tabla F.16: Experimento 3. Configuración PI.

Género	C0	C1	C2	C3	%Contribución
Reggae	8.2 %	4.9 %	6.1 %	5.8 %	25.0 %
Blues	5.9 %	6.7 %	5.5 %	6.9 %	25.0 %
Jazz	6.4 %	5.8 %	6.1 %	6.7 %	25.0 %
Folk	5.9 %	7.0 %	3.8 %	8.3 %	25.0 %
% Ej. Agrupados	26.4 %	24.4 %	21.6 %	27.6 %	

Clasificación general: 29.3 %.

Tabla F.17: Experimento 3. Configuración P.

Género	C0	C1	C2	C3	%Contribución
Reggae	8.2 %	4.9 %	6.1 %	5.8 %	25.0 %
Blues	5.9 %	6.7 %	5.5 %	6.9 %	25.0 %
Jazz	6.4 %	5.8 %	6.1 %	6.7 %	25.0 %
Folk	5.9 %	7.0 %	3.8 %	8.3 %	25.0 %
% Ej. Agrupados	26.5 %	24.4 %	21.5 %	27.6 %	

Clasificación general: 29.3 %.

Tabla F.18: Experimento 3. Configuración I.

Género	C0	C1	C2	C3	%Contribución
Reggae	9.8 %	10.0 %	1.0 %	4.2 %	25.0 %
Blues	9.4 %	9.5 %	1.0 %	5.1 %	25.0 %
Jazz	6.8 %	9.2 %	2.3 %	6.7 %	25.0 %
Folk	7.4 %	9.5 %	1.8 %	6.4 %	25.0 %
% Ej. Agrupados	33.4 %	38.2 %	6.0 %	22.4 %	

Clasificación general: 28.0 %.

Tabla F.19: Experimento 3. Configuración TI.

Género	C0	C1	C2	C3	%Contribución
Blues	11.8 %	7.3 %	2.0 %	3.9 %	25.0 %
Folk	9.0 %	7.5 %	3.6 %	4.9 %	25.0 %
Jazz	8.8 %	6.5 %	4.2 %	5.4 %	25.0 %
Reggae	12.7 %	2.6 %	1.6 %	8.2 %	25.0 %
% Ej. Agrupados	42.3 %	23.8 %	11.4 %	22.5 %	

Clasificación general: 31.8 %.

Tabla F.20: Experimento 3. Configuración T.

Género	C0	C1	C2	C3	%Contribución
Blues	10.8 %	8.0 %	2.2 %	4.0 %	25.0 %
Folk	7.5 %	8.1 %	4.1 %	5.4 %	25.0 %
Jazz	8.0 %	6.7 %	4.5 %	5.8 %	25.0 %
Reggae	12.1 %	2.7 %	1.4 %	8.8 %	25.0 %
% Ej. Agrupados	38.3 %	25.5 %	12.2 %	23.9 %	

Clasificación general: 32.2 %.

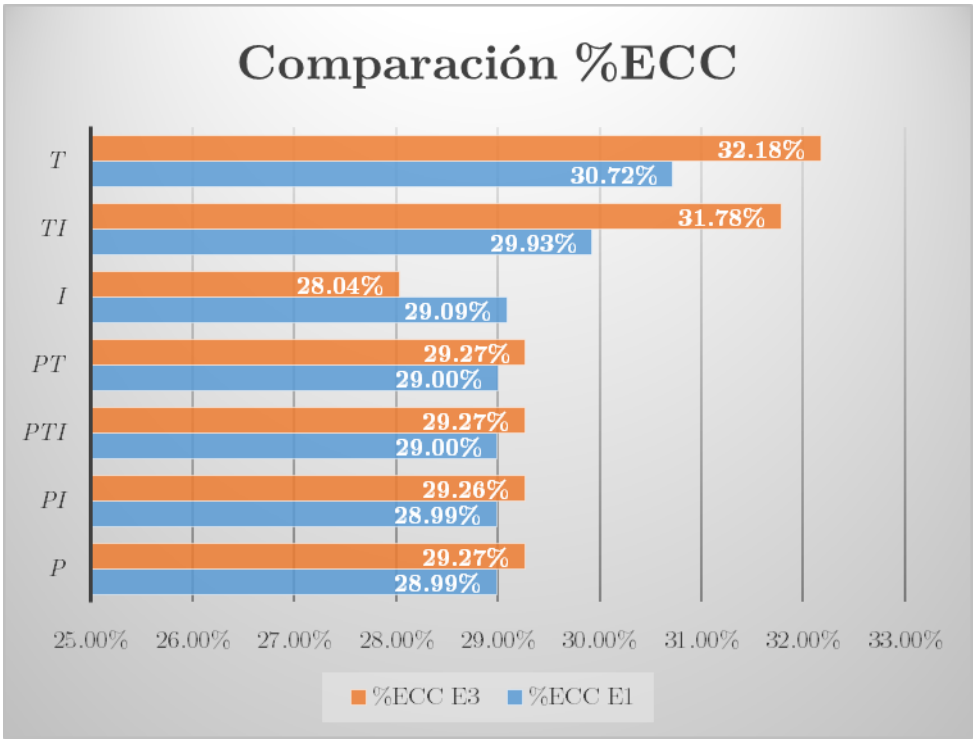

Figura F.3: Comparación de %ECC de los experimentos 1 y 3.

F.4 Experimento 4.

Se utilizó el total de coeficientes del vector del timbre y se balanceo la base tal y como se indicó en el experimento 2 (Tabla F.13).

Los resultados para las configuraciones que no involucran T son los mismos del experimento 3. Por lo tanto sólo se muestran las tablas de resultados para las configuraciones donde aparece T.

Las configuraciones PTI, PT, PI y P agrupan casi la misma cantidad de ejemplos. Al reportarlos como porcentajes con un decimal, el redondeo hace que los resultados sean iguales (Tablas F.16, F.17, F.21 y F.22).

En la figura F.4 se muestran el porcentaje de ejemplos clasificados para el experimento 4 y la comparación con los experimentos 1, 2 y 3.

Tabla F.21: Experimento 4. Configuración PTI.

Género	C0	C1	C2	C3	%Contribución
Reggae	8.2 %	4.9 %	6.1 %	5.8 %	25.0 %
Blues	5.8 %	6.7 %	5.6 %	6.9 %	25.0 %
Jazz	6.4 %	5.8 %	6.1 %	6.7 %	25.0 %
Folk	5.9 %	7.0 %	3.8 %	8.3 %	25.0 %
% Ej. Agrupados	26.4 %	24.4 %	21.6 %	27.6 %	

Clasificación general: 29.3 %.

Tabla F.22: Experimento 4. Configuración PT.

Género	C0	C1	C2	C3	%Contribución
Reggae	8.2 %	4.9 %	6.1 %	5.8 %	25.0 %
Blues	5.9 %	6.7 %	5.5 %	6.9 %	25.0 %
Jazz	6.4 %	5.8 %	6.1 %	6.7 %	25.0 %
Folk	5.9 %	7.0 %	3.8 %	8.3 %	25.0 %
% Ej. Agrupados	26.4 %	24.4 %	21.6 %	27.6 %	

Clasificación general: 29.3 %.

Tabla F.23: Experimento 4. Configuración TI.

Género	C0	C1	C2	C3	%Contribución
Blues	11.7 %	7.3 %	2.0 %	4.0 %	25.0 %
Folk	8.9 %	7.3 %	3.6 %	5.3 %	25.0 %
Jazz	8.8 %	6.3 %	4.3 %	5.5 %	25.0 %
Reggae	12.7 %	2.6 %	1.6 %	8.2 %	25.0 %
% Ej. Agrupados	42.1 %	23.5 %	11.4 %	23.0 %	

Clasificación general: 31.5 %.

Tabla F.24: Experimento 4. Configuración T.

Género	C0	C1	C2	C3	%Contribución
Reggae	12.7 %	2.6 %	1.3 %	8.4 %	25.0 %
Blues	10.2 %	8.2 %	2.2 %	4.4 %	25.0 %
Jazz	7.8 %	6.3 %	4.8 %	6.1 %	25.0 %
Folk	7.0 %	7.6 %	4.2 %	6.2 %	25.0 %
% Ej. Agrupados	37.7 %	24.8 %	12.5 %	25.0 %	

Clasificación general: 31.8 %.

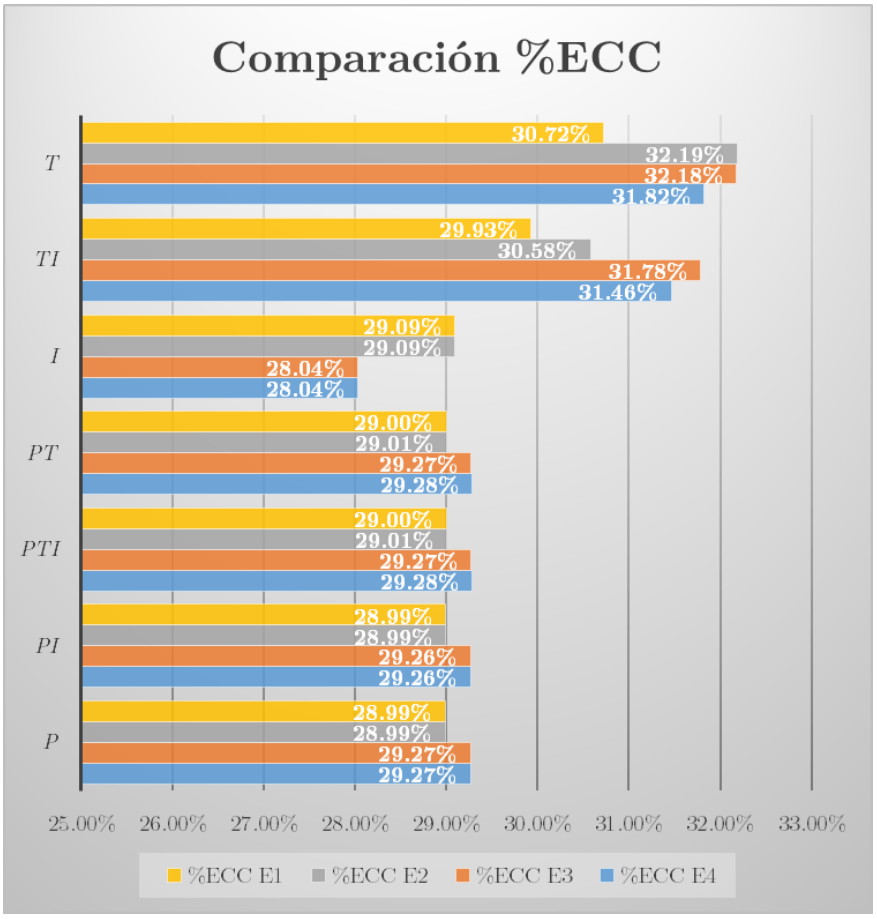

Figura F.4: Comparación de %ECC de los experimentos 1, 2, 3 y 4.

En los experimentos del 5 al 8 se utiliza la Base de datos de la tabla F.25

F.5 Experimento 5.

Se utilizaron los cuatro primeros coeficientes del vector del timbre que están descritos en la documentación (Ver apéndice A.

Las configuraciones PTI, PT, PI y P agrupan casi la misma cantidad de ejemplos. Al reportarlos como porcentajes con un decimal, el redondeo hace que los resultados sean iguales (Tablas F.26, F.27, F.28 y F.29).

En la figura F.5 se muestra el porcentaje de ejemplos clasificados para el experimento 5.

Tabla F.25: Géneros y ejemplos de la BD-2.

Género Musical	No. de Ejemplos	% Ejemplos
Pop Rock	156312	18 %
Electrónica	298316	34 %
Jazz	225089	26 %
Rap	197589	22 %
Total	877306	100 %

Distribución de los ejemplos del conjunto de 200 canciones por géneros.

Tabla F.26: Experimento 5. Configuración PTI.

Género	C0	C1	C2	C3	%Contribución
Rap	4.6 %	7.0 %	5.0 %	5.9 %	22.5 %
Electrónica	6.7 %	11.7 %	7.1 %	8.4 %	34.0 %
Jazz	3.6 %	6.8 %	10.8 %	4.4 %	25.7 %
PopRock	2.5 %	4.3 %	7.0 %	4.0 %	17.8 %
% Ej. Agrupados	17.5 %	29.9 %	29.9 %	22.7 %	

Clasificación general: 31.1 %.

Tabla F.27: Experimento 5. Configuración PT.

Género	C0	C1	C2	C3	%Contribución
Rap	4.6 %	7.0 %	5.0 %	5.9 %	22.5 %
Electrónica	6.7 %	11.7 %	7.1 %	8.4 %	34.0 %
Jazz	3.6 %	6.8 %	10.8 %	4.4 %	25.7 %
Pop Rock	2.5 %	4.3 %	7.0 %	4.0 %	17.8 %
% Ej. Agrupados	17.5 %	29.9 %	30.0 %	22.7 %	

Clasificación general: 31.1 %.

Tabla F.28: Experimento 5. Configuración PI.

Género	C0	C1	C2	C3	%Contribución
Rap	4.6 %	7.0 %	5.0 %	5.9 %	22.5 %
Electrónica	6.7 %	11.7 %	7.1 %	8.4 %	34.0 %
Jazz	3.6 %	6.8 %	10.8 %	4.4 %	25.7 %
Pop Rock	2.5 %	4.3 %	7.0 %	4.0 %	17.8 %
% Ej. Agrupados	17.5 %	29.9 %	29.9 %	22.7 %	

Clasificación general: 31.1 %.

Tabla F.29: Experimento 5. Configuración P.

Género	C0	C1	C2	C3	%Contribución
Rap	4.6 %	7.0 %	5.0 %	5.8 %	22.5 %
Electrónica	6.7 %	11.7 %	7.1 %	8.4 %	34.0 %
Jazz	3.6 %	6.9 %	10.8 %	4.4 %	25.7 %
Pop Rock	2.5 %	4.3 %	7.0 %	3.9 %	17.8 %
% Ej. Agrupados	17.5 %	29.9 %	29.9 %	22.7 %	

Clasificación general: 31.1 %.

Tabla F.30: Experimento 5. Configuración I.

Género	C0	C1	C2	C3	%Contribución
Rap	0.8 %	3.3 %	8.7 %	9.8 %	22.5 %
Jazz	2.5 %	7.7 %	9.8 %	5.6 %	25.7 %
Electrónica	2.0 %	7.4 %	13.4 %	11.2 %	34.0 %
Pop Rock	0.7 %	2.0 %	5.4 %	9.8 %	17.8 %
% Ej. Agrupados	5.9 %	20.4 %	37.2 %	36.5 %	

Clasificación general: 31.7 %.

Tabla F.31: Experimento 5. Configuración TI.

Género	C0	C1	C2	C3	%Contribución
Pop Rock	1.0 %	3.0 %	2.5 %	11.3 %	17.8 %
Jazz	4.4 %	8.2 %	5.2 %	7.9 %	25.7 %
Electrónica	3.5 %	5.9 %	11.2 %	13.4 %	34.0 %
Rap	1.2 %	3.4 %	6.1 %	11.9 %	22.5 %
% Ej. Agrupados	10.1 %	20.4 %	24.9 %	44.5 %	

Clasificación general: 32.2 %.

Tabla F.32: Experimento 5. Configuración T.

Género	C0	C1	C2	C3	%Contribución
Jazz	7.9 %	6.7 %	5.2 %	5.8 %	25.7 %
Pop Rock	2.1 %	8.3 %	2.6 %	4.7 %	17.8 %
Electrónica	6.5 %	7.7 %	12.0 %	7.9 %	34.0 %
Rap	2.5 %	6.4 %	6.4 %	7.2 %	22.5 %
% Ej. Agrupados	19.0 %	29.2 %	26.1 %	25.6 %	

Clasificación general: 35.4 %.

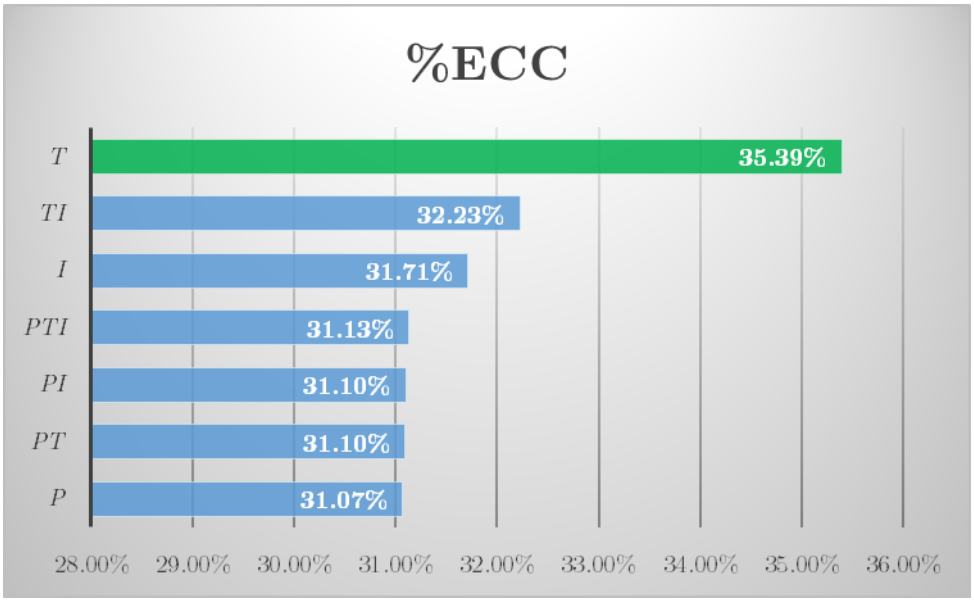

Figura F.5: %ECC:Porcentaje de ejemplos clasificados correctamente del experimento 5.

F.6 Experimento 6.

Se utilizó el total de coeficientes del vector del timbre.

Los resultados para las configuraciones que no involucran T son los mismos del experimento 1. Por lo tanto sólo se muestran las tablas de resultados para las configuraciones donde aparece T.

Las configuraciones PTI, PT, PI y P agrupan casi la misma cantidad de ejemplos. Al reportarlos como porcentajes con un decimal, el redondeo hace que los resultados sean iguales (Tablas F.28, F.29, F.33 y F.34).

En la figura F.6 se muestran el porcentaje de ejemplos clasificados para el experimento 6 y la comparación con el experimento 5.

Tabla F.33: Experimento 6. Configuración PTI.

Género	C0	C1	C2	C3	%Contribución
Rap	4.6 %	7.0 %	5.0 %	5.9 %	22.5 %
Electrónica	6.7 %	11.7 %	7.1 %	8.4 %	34.0 %
Jazz	3.6 %	6.8 %	10.8 %	4.4 %	25.7 %
Pop Rock	2.5 %	4.3 %	7.0 %	4.0 %	17.8 %
% Ej. Agrupados	17.5 %	29.8 %	29.9 %	22.7 %	

Clasificación general: 31.2 %.

Tabla F.34: Experimento 6. Configuración PT.

Género	C0	C1	C2	C3	%Contribución
Rap	4.6 %	7.0 %	5.0 %	5.9 %	22.5 %
Electrónica	6.7 %	11.7 %	7.1 %	8.4 %	34.0 %
Jazz	3.6 %	6.8 %	10.8 %	4.4 %	25.7 %
Pop Rock	2.5 %	4.3 %	7.0 %	4.0 %	17.8 %
% Ej. Agrupados	17.5 %	29.9 %	30.0 %	22.7 %	

Clasificación general: 31.1 %.

Tabla F.35: Experimento 6. Configuración TI.

Género	C0	C1	C2	C3	%Contribución
Pop Rock	0.9 %	3.0 %	2.6 %	11.3 %	17.8 %
Jazz	4.1 %	9.0 %	4.6 %	8.0 %	25.7 %
Electrónica	3.0 %	5.7 %	11.3 %	14.0 %	34.0 %
Rap	1.1 %	3.4 %	5.8 %	12.2 %	22.5 %
% Ej. Agrupados	9.1 %	21.0 %	24.3 %	45.6 %	

Clasificación general: 33.4 %.

Tabla F.36: Experimento 6. Configuración T.

Género	C0	C1	C2	C3	%Contribución
Pop Rock	1.0 %	3.8 %	2.6 %	10.5 %	17.8 %
Jazz	4.6 %	9.1 %	4.8 %	7.0 %	25.7 %
Electrónica	2.8 %	6.1 %	11.1 %	14.0 %	34.0 %
Rap	1.0 %	3.7 %	5.6 %	12.2 %	22.5 %
% Ej. Agrupados	9.5 %	22.8 %	24.0 %	43.7 %	

Clasificación general: 33.4 %.

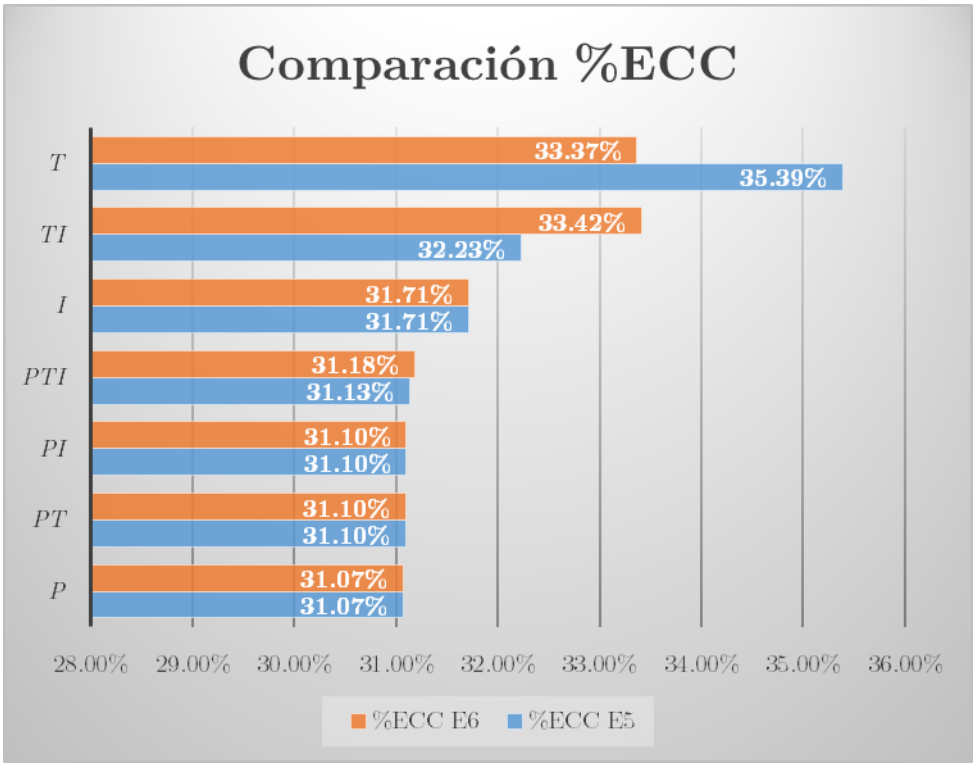

Figura F.6: Comparación de %ECC de los experimentos 5 y 6.

F.7 Experimento 7.

Se utilizaron los cuatro primeros coeficientes del vector del timbre y se balanceo la base igualando la cantidad de ejemplos por género, al número de ejemplos del género con menor cantidad de la BD-1, en este caso Folk (Tabla F.13).

Las configuraciones PTI, PT, PI y P agrupan casi la misma cantidad de ejemplos. Al reportarlos como porcentajes con un decimal, el redondeo hace que los resultados sean iguales (Tablas F.38, F.39, F.40 y F.41).

En la figura F.7 se muestran el porcentaje de ejemplos clasificados para el experimento 7 y la comparación con el experimento 5.

Tabla F.37: Géneros y ejemplos de la BD-2 balanceada.

Género Musical	No. de Ejemplos	% Ejemplos
Pop Rock	149231	25 %
Electrónica	149231	25 %
Rap	149231	25 %
Jazz	149231	25 %
Total	596924	100 %

Distribución de los ejemplos por géneros.

Tabla F.38: Experimento 7. Configuración PTI.

Género	C0	C1	C2	C3	%Contribución
Jazz	10.6 %	4.1 %	3.7 %	6.6 %	25.0 %
Pop Rock	9.6 %	5.6 %	3.7 %	6.0 %	25.0 %
Rap	5.5 %	6.6 %	5.3 %	7.6 %	25.0 %
Electrónica	5.2 %	6.2 %	5.3 %	8.4 %	25.0 %
% Ej. Agrupados	30.9 %	22.6 %	18.0 %	28.6 %	

Clasificación general: 29.8 %.

Tabla F.39: Experimento 7. Configuración PT.

Género	C0	C1	C2	C3	%Contribución
Jazz	10.5 %	4.1 %	3.8 %	6.6 %	25.0 %
Pop Rock %	9.7 %	5.6 %	3.7 %	6.0 %	25.0 %
Rap %	5.5 %	6.6 %	5.3 %	7.6 %	25.0 %
Electrónica %	5.1 %	6.2 %	5.3 %	8.4 %	25.0 %
% Ej. Agrupados	30.9 %	22.5 %	18.0 %	28.6 %	

Clasificación general: 29.8 %.

Tabla F.40: Experimento 7. Configuración PI.

Género	C0	C1	C2	C3	%Contribución
Jazz	10.5 %	4.1 %	3.8 %	6.6 %	25.0 %
Pop Rock	9.7 %	5.6 %	3.7 %	6.0 %	25.0 %
Rap	5.5 %	6.6 %	5.3 %	7.6 %	25.0 %
Electrónica	5.1 %	6.2 %	5.3 %	8.3 %	25.0 %
% Ej. Agrupados	30.9 %	22.5 %	18.0 %	28.6 %	

Clasificación general: 29.8 %.

Tabla F.41: Experimento 7. Configuración P.

Género	C0	C1	C2	C3	%Contribución
Jazz	10.5 %	4.1 %	3.8 %	6.6 %	25.0 %
Pop Rock	9.7 %	5.6 %	3.7 %	6.0 %	25.0 %
Rap	5.5 %	6.6 %	5.3 %	7.6 %	25.0 %
Electrónica	5.1 %	6.2 %	5.3 %	8.4 %	25.0 %
% Ej. Agrupados	30.9 %	22.5 %	18.0 %	28.6 %	

Clasificación general: 29.7 %.

Tabla F.42: Experimento 7. Configuración I.

Género	C0	C1	C2	C3	%Contribución
Rap	10.0 %	1.0 %	10.2 %	3.8 %	25.0 %
Electrónica	9.7 %	1.9 %	7.0 %	6.3 %	25.0 %
Pop Rock	7.8 %	0.9 %	13.4 %	2.8 %	25.0 %
Jazz	9.6 %	2.8 %	4.8 %	7.7 %	25.0 %
% Ej. Agrupados	37.3 %	6.5 %	35.5 %	20.7 %	

Clasificación general: 33.1 %.

Tabla F.43: Experimento 7. Configuración TI.

Género	C0	C1	C2	C3	%Contribución
Rap	6.4 %	1.3 %	12.5 %	4.7 %	25.0 %
Electrónica	7.3 %	3.3 %	8.6 %	5.8 %	25.0 %
Pop Rock	3.4 %	1.4 %	15.3 %	4.9 %	25.0 %
Jazz	4.9 %	5.0 %	6.9 %	8.2 %	25.0 %
% Ej. Agrupados	22.1 %	11.0 %	43.3 %	23.7 %	

Clasificación general: 33.2 %.

Tabla F.44: Experimento 7. Configuración T.

Género	C0	C1	C2	C3	%Contribución
Electrónica	8.8 %	5.0 %	5.1 %	6.1 %	25.0 %
Pop Rock	5.1 %	10.5 %	6.4 %	2.9 %	25.0 %
Rap	8.2 %	6.2 %	7.8 %	2.9 %	25.0 %
Jazz	5.2 %	6.7 %	5.4 %	7.8 %	25.0 %
% Ej. Agrupados	27.3 %	28.3 %	24.7 %	19.7 %	

Clasificación general: 34.8 %.

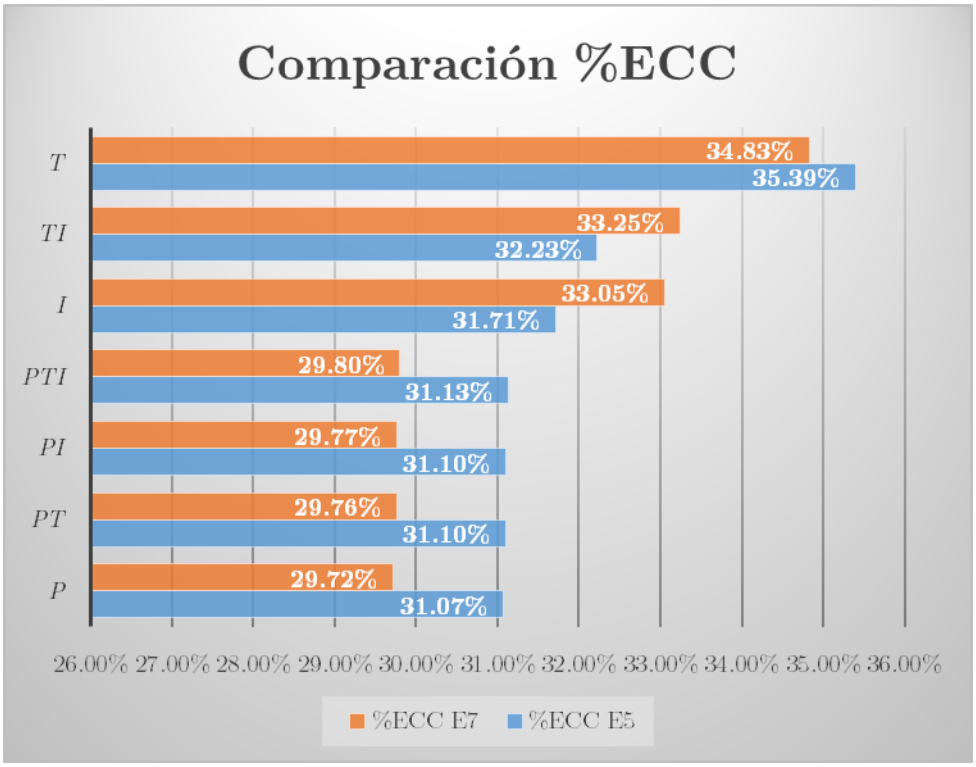

Figura F.7: Comparación de %ECC de los experimentos 5 y 7.

F.8 Experimento 8.

Se utilizó el total de coeficientes del vector del timbre y se balanceo la base tal y como se indicó en el experimento 6 (Tabla F.37).

Los resultados para las configuraciones que no involucran T son los mismos del experimento 7. Por lo tanto sólo se muestran las tablas de resultados para las configuraciones donde aparece T.

Las configuraciones PTI, PT, PI y P agrupan casi la misma cantidad de ejemplos. Al reportarlos como porcentajes con un decimal, el redondeo hace que los resultados sean iguales (Tablas F.40, F.41, F.45 y F.46).
En la figura F.8 se muestran el porcentaje de ejemplos clasificados para el experimento 8 y la comparación con los experimentos 5, 6 y 7.

Tabla F.45: Experimento 8. Configuración PTI.

Género	C0	C1	C2	C3	%Contribución
Jazz	10.6 %	4.1 %	3.7 %	6.6 %	25.0 %
Pop Rock	9.6 %	5.7 %	3.7 %	6.0 %	25.0 %
Rap	5.5 %	6.7 %	5.3 %	7.6 %	25.0 %
Electrónica	5.2 %	6.2 %	5.3 %	8.4 %	25.0 %
% Ej. Agrupados	30.8 %	22.6 %	18.0 %	28.6 %	

Clasificación general: 29.9 %

Tabla F.46: Experimento 8. Configuración PT.

Género	C0	C1	C2	C3	%Contribución
Jazz	10.6 %	4.1 %	3.7 %	6.6 %	25.0 %
Pop Rock	9.7 %	5.6 %	3.7 %	6.0 %	25.0 %
Rap	5.5 %	6.6 %	5.3 %	7.6 %	25.0 %
Electrónica	5.2 %	6.2 %	5.3 %	8.4 %	25.0 %
% Ej. Agrupados	30.9 %	22.6 %	18.0 %	28.5 %	

Clasificación general: 29.8 %.

Tabla F.47: Experimento 8. Configuración TI.

Género	C0	C1	C2	C3	%Contribución
Rap	6.9 %	3.4 %	13.5 %	1.3 %	25.0 %
Jazz	5.6 %	7.8 %	7.2 %	4.4 %	25.0 %
Electrónica	4.1 %	3.9 %	15.7 %	1.3 %	25.0 %
Pop Rock	7.1 %	5.5 %	9.8 %	2.7 %	25.0 %
% Ej. Agrupados	23.6 %	20.6 %	46.1 %	9.7 %	

Clasificación general: 33.0 %.

Tabla F.48: Experimento 8. Configuración T.

Género	C0	C1	C2	C3	%Contribución
Rap	5.7 %	4.9 %	13.4 %	1.1 %	25.0 %
Electrónica	5.9 %	6.6 %	9.8 %	2.7 %	25.0 %
Pop Rock	3.6 %	5.5 %	14.5 %	1.3 %	25.0 %
Jazz	5.0 %	8.5 %	6.6 %	4.9 %	25.0 %
% Ej. Agrupados	20.2 %	25.5 %	44.3 %	10.0 %	

Clasificación general: 31.6 %.

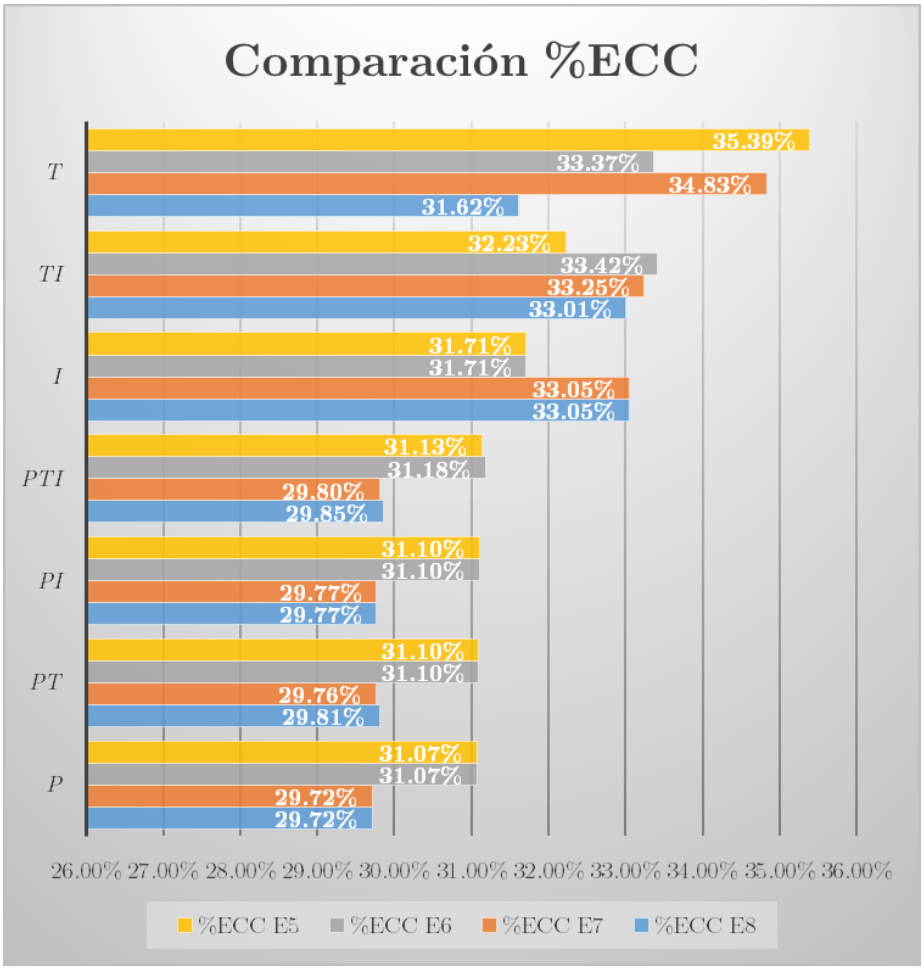

Figura F.8: Comparación de %ECC de los experimentos 5, 6, 7 y 8.

F.9 Experimentos con Promedios.

Los siguientes experimentos se realizaron para observar si se podría acelerar el proceso utilizando sólo el promedio por columnas del conjunto de segmentos de las características de pitch, intensidad y timbre. Además para observar si el comportamiento del conjunto de datos y las configuraciones prevalecen con esta nueva presentación de los ejemplos.

Las bases de datos nuevamente utilizan 200 canciones por género de forma que ahora desde un principio la base de datos ya esta balanceada al no utilizar uno a uno los segmentos sino el promedio.

Se utilizaron las características tono o pitch (P), timbre (T) e intensidad (I).
Se utilizó un conjunto reducido de configuraciones para observar cómo las características seleccionadas afectaban el agrupamiento.

$$Configuraciones = \{P, T, I, PTI\}$$

En los experimentos 9 y 10 se utilizaron los géneros de la tabla 4.1.

En los experimentos 11 y 12 se utilizaron los géneros de la tabla 4.5.

En los experimentos 13 y 14 se utilizaron los géneros: *Clasica*, *Electrónica*, *Blues* y *Rap* cada uno con 200 canciones.

F.9.1 Experimento 9.

Se utilizaron los cuatro primeros coeficientes del vector del timbre.

Los Resultados se muestran en las tablas F.49, F.50, F.51 y F.52.

Tabla F.49: Experimento 9. Configuración PTI.

Género	C0	C1	C2	C3	%Contribución
Jazz	9.6 %	3.4 %	7.3 %	4.8 %	25.0 %
Blues	9.0 %	5.8 %	4.8 %	5.5 %	25.0 %
Folk	3.8 %	8.3 %	9.3 %	3.8 %	25.0 %
Reggae	11.5 %	1.1 %	0.4 %	12.0 %	25.0 %
% Ej. Agrupados	33.9 %	18.5 %	21.6 %	26.0 %	

Clasificación general: 36.6 %.

Tabla F.50: Experimento 9. Configuración P.

Género	C0	C1	C2	C3	%Contribución
Reggae	16.5 %	2.1 %	3.9 %	2.5 %	25.0 %
Folk	3.5 %	7.5 %	7.8 %	6.3 %	25.0 %
Jazz	11.0 %	3.8 %	8.0 %	2.3 %	25.0 %
Blues	8.1 %	5.1 %	5.3 %	6.5 %	25.0 %
% Ej. Agrupados	39.1 %	18.5 %	24.9 %	17.5 %	

Clasificación general: 38.5 %.

Tabla F.51: Experimento 9. Configuración I.

Género	C0	C1	C2	C3	%Contribución
Jazz	5.0 %	3.4 %	7.6 %	9.0 %	25.0 %
Blues	3.5 %	8.3 %	5.3 %	8.0 %	25.0 %
Folk	3.6 %	4.0 %	9.3 %	8.1 %	25.0 %
Reggae	0.8 %	9.5 %	5.4 %	9.4 %	25.0 %
% Ej. Agrupados	12.9 %	25.1 %	27.5 %	34.5 %	

Clasificación general: 31.9 %.

Tabla F.52: Experimento 9. Configuración T.

Género	C0	C1	C2	C3	%Contribución
Jazz	6.5 %	6.9 %	8.6 %	3.0 %	25.0 %
Reggae	1.3 %	17.5 %	1.4 %	4.9 %	25.0 %
Folk	6.0 %	5.1 %	10.5 %	3.4 %	25.0 %
Blues	6.8 %	4.5 %	4.1 %	9.6 %	25.0 %
% Ej. Agrupados	20.5 %	34.0 %	24.6 %	20.9 %	

Clasificación general: 44.1 %.

F.9.2 Experimento 10.

Se utilizó el total de coeficientes del vector del timbre.

Los resultados para las configuraciones que no involucran T son los mismos del experimento 9. Por lo tanto sólo se muestran las tablas de resultados para las configuraciones donde aparece T.

Los Resultados se muestran en las tablas F.53 y F.54.

En la figura F.9 se muestran el porcentaje de ejemplos clasificados para el experimento 10 y la comparación con el experimento 9.

Tabla F.53: Experimento 10. Configuración PTI.

Género	C0	C1	C2	C3	%Contribución
Blues	5.9 %	8.5 %	2.6 %	8.0 %	25.0 %
Folk	3.1 %	11.6 %	7.5 %	2.8 %	25.0 %
Jazz	5.5 %	5.0 %	6.1 %	8.4 %	25.0 %
Reggae	3.4 %	2.3 %	0.0 %	19.4 %	25.0 %
% Ej. Agrupados	17.9 %	27.4 %	16.3 %	38.5 %	

Clasificación general: 43.0 %.

Tabla F.54: Experimento 10. Configuración T.

Género	C0	C1	C2	C3	%Contribución
Folk	8.9 %	11.1 %	3.5 %	1.5 %	25.0 %
Jazz	7.3 %	10.9 %	4.3 %	2.6 %	25.0 %
Blues	2.5 %	12.4 %	6.8 %	3.4 %	25.0 %
Reggae	0.5 %	5.4 %	0.3 %	18.9 %	25.0 %
% Ej. Agrupados	19.1 %	39.8 %	14.8 %	26.4 %	

Clasificación general: 45.4 %.

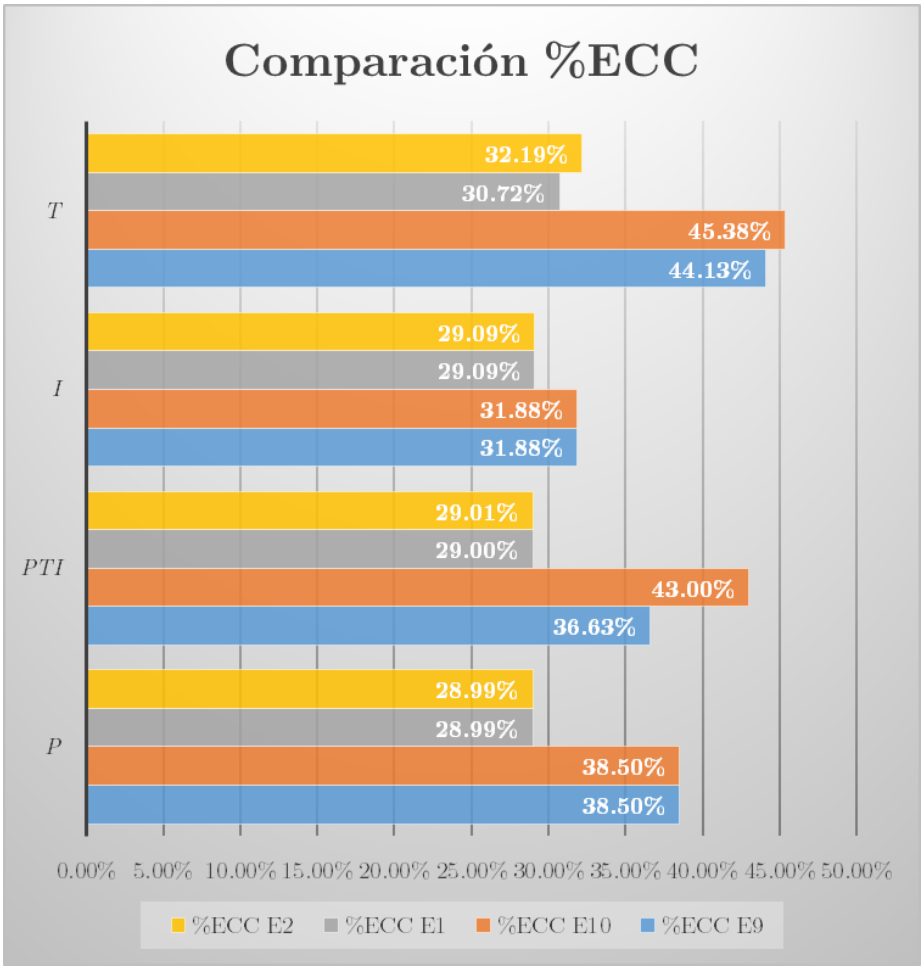

Figura F.9: Comparación de %ECC de los experimentos 1, 2, 9 y 10.

F.9.3 Experimento 11.

Se utilizaron los cuatro primeros coeficientes del vector del timbre. Los Resultados se muestran en las tablas F.55, F.56, F.57 y F.58.

Tabla F.55: Experimento 11. Configuración PTI.

Género	C0	C1	C2	C3	%Contribución
Jazz	9.1 %	2.0 %	4.6 %	9.3 %	25.0 %
Pop Rock	3.6 %	11.4 %	2.3 %	7.8 %	25.0 %
Electrónica	2.3 %	9.1 %	10.0 %	3.6 %	25.0 %
Rap	1.0 %	9.9 %	6.1 %	8.0 %	25.0 %
% Ej. Agrupados	16.0 %	32.4 %	23.0 %	28.6 %	

Clasificación general: 38.5 %.

Tabla F.56: Experimento 11. Configuración P.

Género	C0	C1	C2	C3	%Contribución
Jazz	10.3 %	5.0 %	1.5 %	8.3 %	25.0 %
Electrónica	2.8 %	9.3 %	8.1 %	4.9 %	25.0 %
Pop Rock	6.1 %	3.5 %	9.5 %	5.9 %	25.0 %
Rap	1.5 %	6.0 %	8.4 %	9.1 %	25.0 %
% Ej. Agrupados	20.6 %	23.8 %	27.5 %	28.1 %	

Clasificación general: 38.1 %.

Tabla F.57: Experimento 11. Configuración I.

Género	C0	C1	C2	C3	%Contribución
Electrónica	9.1 %	9.0 %	2.3 %	4.6 %	25.0 %
Rap	6.3 %	12.4 %	1.1 %	5.3 %	25.0 %
Jazz	11.1 %	6.9 %	6.0 %	1.0 %	25.0 %
Pop Rock	4.3 %	7.9 %	0.9 %	12.0 %	25.0 %
% Ej. Agrupados	30.8 %	36.1 %	10.3 %	22.9 %	

Clasificación general: 39.5 %.

Tabla F.58: Experimento 11. Configuración T.

Género	C0	C1	C2	C3	%Contribución
Jazz	13.0 %	0.8 %	7.0 %	4.3 %	25.0 %
Electrónica	4.6 %	8.0 %	8.5 %	3.9 %	25.0 %
Rap	1.8 %	1.8 %	14.9 %	6.6 %	25.0 %
Pop Rock	3.5 %	0.3 %	5.1 %	16.1 %	25.0 %
% Ej. Agrupados	22.9 %	10.8 %	35.5 %	30.9 %	

Clasificación general: 52.0 %.

F.9.4 Experimento 12.

Se utilizó el total de coeficientes del vector del timbre.

Los resultados para las configuraciones que no involucran T son los mismos del experimento 11. Por lo tanto sólo se muestran las tablas de resultados para las configuraciones donde aparece T.

Los Resultados se muestran en las tablas F.59 y F.60.

En la figura F.10 se muestran el porcentaje de ejemplos clasificados para el experimento 11 y la comparación con el experimento 12.

Tabla F.59: Experimento 12. Configuración PTI.

Género	C0	C1	C2	C3	%Contribución
Jazz	9.9 %	1.3 %	9.9 %	4.0 %	25.0 %
Pop Rock	3.1 %	11.6 %	8.1 %	2.1 %	25.0 %
Electrónica	2.8 %	7.4 %	8.0 %	6.9 %	25.0 %
Rap	0.9 %	6.0 %	6.9 %	11.3 %	25.0 %
% Ej. Agrupados	16.6 %	26.3 %	32.9 %	24.3 %	

Clasificación general: 40.8 %.

Tabla F.60: Experimento 12. Configuración T.

Género	C0	C1	C2	C3	%Contribución
Jazz	9.1 %	2.9 %	9.9 %	3.1 %	25.0 %
Rap	0.8 %	16.4 %	3.8 %	4.1 %	25.0 %
Electrónica	2.6 %	5.4 %	8.8 %	8.3 %	25.0 %
Pop Rock	1.5 %	1.1 %	7.6 %	14.8 %	25.0 %
% Ej. Agrupados	14.0 %	25.8 %	30.0 %	30.3 %	

Clasificación general: 49.0 %.

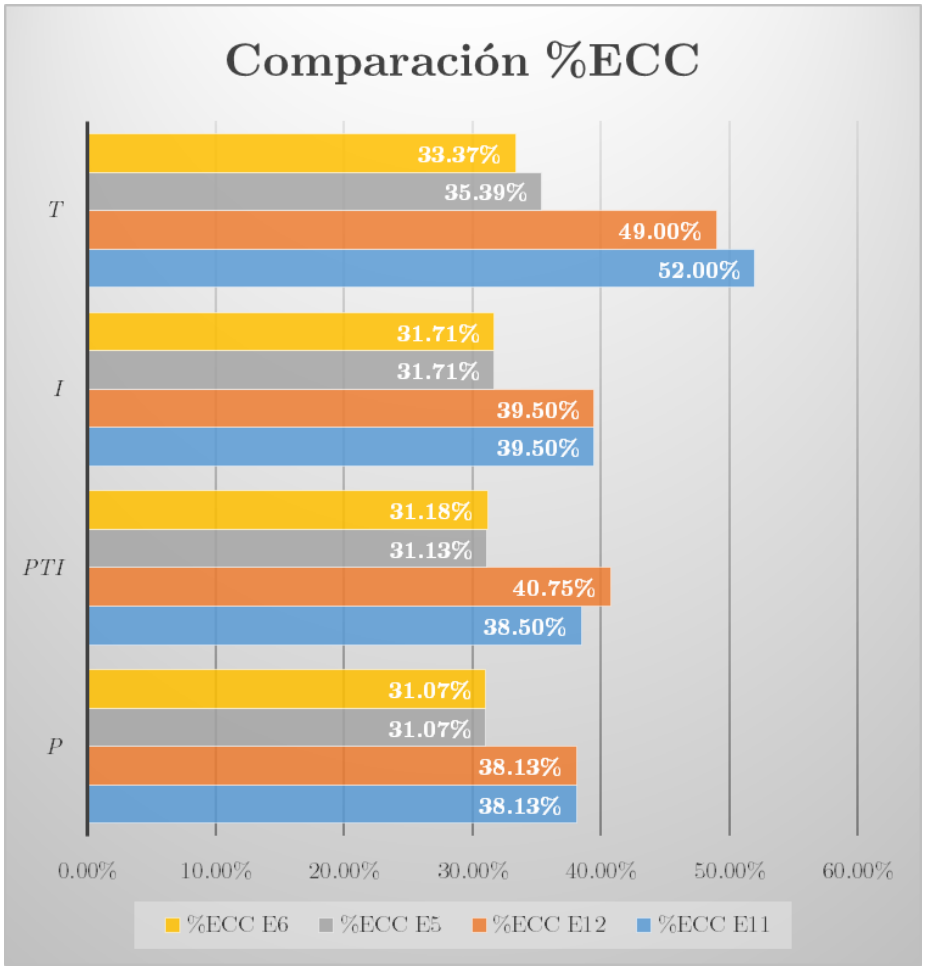

Figura F.10: Comparación de %ECC de los experimentos 5, 6, 11 y 12.

F.9.5 Experimento 13.

Se utilizaron los cuatro primeros coeficientes del vector del timbre. Los Resultados se muestran en las tablas F.61, F.62, F.63 y F.64.

Tabla F.61: Experimento 13. Configuración PTI.

Género	C0	C1	C2	C3	%Contribución
Clásica	15.1 %	3.3 %	5.6 %	1.0 %	25.0 %
Electrónica	1.5 %	10.1 %	2.4 %	11.0 %	25.0 %
Blues	4.4 %	12.3 %	6.0 %	2.4 %	25.0 %
Rap	0.5 %	10.6 %	1.0 %	12.9 %	25.0 %
% Ej. Agrupados	21.5 %	36.3 %	15.0 %	27.3 %	

Clasificación general: 44.1 %.

Tabla F.62: Experimento 13. Configuración P.

Género	C0	C1	C2	C3	%Contribución
Clásica	15.0 %	2.3 %	6.0 %	1.8 %	25.0 %
Electrónica	2.4 %	10.6 %	2.0 %	10.0 %	25.0 %
Blues	8.0 %	5.8 %	6.8 %	4.5 %	25.0 %
Rap	1.1 %	7.9 %	2.6 %	13.4 %	25.0 %
% Ej. Agrupados	26.5 %	26.5 %	17.4 %	29.6 %	

Clasificación general: 45.8 %.

Tabla F.63: Experimento 13. Configuración I.

Género	C0	C1	C2	C3	%Contribución
Clásica	6.9 %	2.9 %	7.6 %	7.6 %	25.0 %
Rap	0.4 %	14.6 %	1.9 %	8.1 %	25.0 %
Blues	0.6 %	9.6 %	6.0 %	8.8 %	25.0 %
Electrónica	0.9 %	9.5 %	3.3 %	11.4 %	25.0 %
% Ej. Agrupados	8.8 %	36.6 %	18.8 %	35.9 %	

Clasificación general: 38.9 %.

Tabla F.64: Experimento 13. Configuración T.

Género	C0	C1	C2	C3	%Contribución
Blues	9.6 %	11.6 %	3.1 %	0.6 %	25.0 %
Rap	1.1 %	20.9 %	0.6 %	2.4 %	25.0 %
Clásica	6.6 %	5.5 %	12.9 %	0.0 %	25.0 %
Electrónica	2.1 %	11.9 %	2.5 %	8.5 %	25.0 %
% Ej. Agrupados	19.5 %	49.9 %	19.1 %	11.5 %	

Clasificación general: 51.9 %.

F.9.6 Experimento 14.

Se utilizó el total de coeficientes del vector del timbre.

Los resultados para las configuraciones que no involucran T son los mismos del experimento 13. Por lo tanto sólo se muestran las tablas de resultados para las configuraciones donde aparece T.

Los Resultados se muestran en las tablas F.65 y F.66.

En la figura F.11 se muestran el porcentaje de ejemplos clasificados para el experimento 14 y la comparación con el experimento 13.

Tabla F.65: Experimento 14. Configuración PTI.

Género	C0	C1	C2	C3	%Contribución
Clásica	16.3 %	5.6 %	2.1 %	1.0 %	25.0 %
Electrónica	2.3 %	8.9 %	0.3 %	13.6 %	25.0 %
Blues	3.3 %	13.5 %	6.1 %	2.1 %	25.0 %
Rap	0.6 %	6.6 %	1.1 %	16.6 %	25.0 %
% Ej. Agrupados	22.4 %	34.6 %	9.6 %	33.4 %	

Clasificación general: 47.9 %.

Tabla F.66: Experimento 14. Configuración T.

Género	C0	C1	C2	C3	%Contribución
Clásica	11.6 %	8.0 %	4.5 %	0.9 %	25.0 %
Electrónica	1.8 %	14.0 %	0.4 %	8.9 %	25.0 %
Blues	1.3 %	15.0 %	7.3 %	1.5 %	25.0 %
Rap	0.4 %	6.8 %	1.3 %	16.6 %	25.0 %
% Ej. Agrupados	15.0 %	43.8 %	13.4 %	27.9 %	

Clasificación general: 49.5 %.

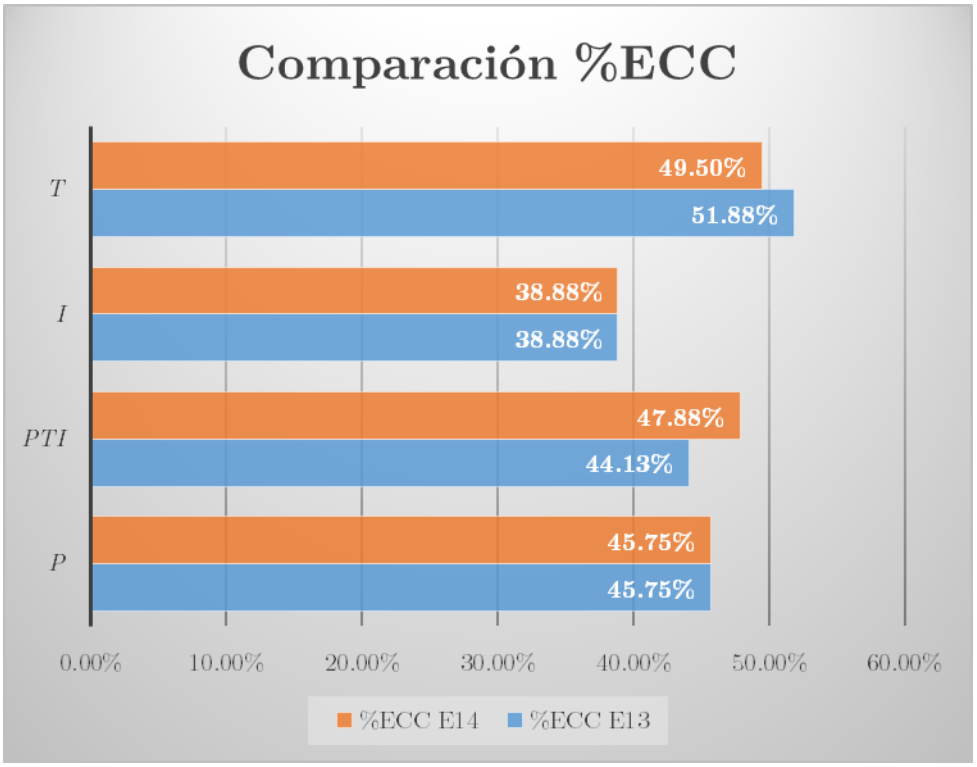

Figura F.11: Comparación de %ECC de los experimentos 13 y 14.

F.10 DBN Experimentos con T_4.

En la tabla F.67 se muestran los resultados obtenidos utilizando las redes de creencia profunda. La primer columna muestra la distribución de los nodos dentro de la red indicando el número de nodos utilizados por capa. La siguientes dos columnas indican la taza de error obtenida para el conjunto de prueba y para el de entrenamiento respectivamente. La última indica el número máximo de iteraciones a utilizar.

Se fijo el número de capas ocultas a 3, la cantidad de nodos de la capa de entrada a 4 correspondiendo con el número de variables utilizadas y la cantidad de nodos de la capa de salida a 6 correspondiendo con el número de géneros utilizados.

La heurística utilizada para depurar la red consistió en asignar primero de forma aleatoria la cantidad de nodos de las capas ocultas con un número entre 0 y 24 (10 16 0).
Después se incremento o decremento de dos en dos el número de nodos de cada una de las capas ocultas empezando por la más cercana a la salida de la red. Se fijó el número de nodos de la capa que correspondió con el menor **ErrorRate Test**.

Tabla F.67: Tasas de Error Aprendizaje Profundo.

Nodos X Capa	ErrorRate Test	ErrorRate Train	Máx. Iteraciones
[4 10 12 14 6]	0.676461	0.681109	150
[4 12 12 14 6]	0.676686	0.677522	150
[4 10 16 14 6]	0.680778	0.681794	150
[4 10 10 14 6]	0.681703	0.682422	150
[4 10 16 6 6]	0.681887	0.680602	150
[4 8 12 14 6]	0.682634	0.684384	150
[4 16 12 14 6]	0.684182	0.681373	150
[4 10 16 4 6]	0.685328	0.681118	150
[4 10 16 10 6]	0.685609	0.683664	150
[4 6 12 14 6]	0.685982	0.692306	150
[4 14 12 14 6]	0.685987	0.679064	150
[4 10 14 14 6]	0.686034	0.682465	150
[4 10 16 2 6]	0.686954	0.698737	150
[4 10 22 14 6]	0.687001	0.683827	150
[4 10 18 14 6]	0.687206	0.682102	150
[4 10 16 8 6]	0.689528	0.683362	150
[4 10 20 14 6]	0.689618	0.682519	150
[4 10 16 12 6]	0.695382	0.684825	150
[4 4 12 14 6]	0.695504	0.701468	150
[4 10 16 12 6]	0.719933	0.720401	500
[4 10 16 6]	0.764623	0.807349	1000

Tasas de Error Aprendizaje Profundo.