



Pontificia Universidad Católica de Chile
Escuela de Ingeniería
Departamento de Ingeniería Eléctrica

IEE3724 Reconocimiento de Patrones

Proyecto - Entrega 1

Constanza Villanueva - Rodrigo González

Mayo 15, 2014

Reconocimiento de Géneros Musicales

- Objetivos

La música es un elemento que se encuentra en gran cantidad en la web hoy en día. La clasificación de esta según su género es un herramienta muy utilizada para algoritmos de búsqueda y categorización.

- ▶ Se desea desarrollar, en vista de lo anterior, un programa que permita identificar el género musical de una pista de audio, de un conjunto preseleccionado de clases.
- ▶ Se debe elegir un conjunto apropiado de *Feature Vectors* que permita analizar los distintos componentes que definen un género de otro.
- ▶ La elección del género particular de cada audio es muy suceptible a ambigüedades, ya que una canción puede tener mezcla de géneros y el género principal elegido en los labels utilizados puede ser por elección personal. En vista de esto hay que tener en cuenta posible subclasificaciones como método de extender la clasificación.

Esquema del Programa a Implementar

- Para implementar el programa se utilizará Matlab.
- La entrada al programa serán las pistas de audio en conjunto con los *labels* de las clases, tanto para los sets de entrenamiento como para los de validación y testeo.
- Se analizará el uso de distintos tipos de *Features* en distintas combinaciones, además de comparar métodos de clasificación.
- La salida del programa entregará las matrices de confusión de las distintas pruebas.

Restricciones

- **Formato de Audio:** Se trabajará con audio en los siguientes posibles formatos: MP3, WAV y AU. En general, se buscan formatos que permitan una buena calidad de audio y que no sea necesaria una conversión extra para su trabajo en Matlab. En ese sentido no se trabajará con audios monofónicos como los en formato MIDI.
- **Duración de las pistas:** El largo de las pistas puede ser cualquiera, pudiendo utilizarse tanto el total de la canción, lo que permitirá un mayor análisis de ventanas, como trozos de las pistas. La restricción ira en que la mínima duración a trabajar es de 10 segundos.
- **Clases:** El programa solo será capaz de analizar géneros musicales de un conjunto determinado de clases, cualquier otro género será clasificado según su aproximación más cercana a los géneros disponibles.

Metodologías Actuales

Existen en la literatura actual diversos tipos de características para el análisis de señales de audio, entre estos grupos destacan:

- **Características Temporales:**

Se basan en el estudio de la señal en el dominio del tiempo como son los cálculos de energía y razón de cruces por cero.

- **Características Espectrales:**

Corresponde al grupo más desarrollado de características donde se analiza el contenido en frecuencia de la señal. Por un lado, existen características simples, como el cálculo del centroide espectral, la uniformidad del espectro, su tasa de crestas o bien el flujo espectral.

Por otro lado, existen otras más complejas, como veremos adelante, correspondientes a los cepstrum, los que permiten análisis del contenido armónico.

- **Características Rítmicas:**

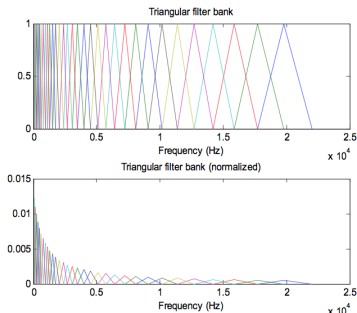
Una mezcla de ambos aspectos anteriores con el fin de determinar detalles rítmicos, como son el tempo o *beat histograms*.

Metodologías Actuales

- **“Music Genre Classification Algorithm Based on Dynamic Frame Analysis and Support Vector Machine”, Shih-Hao *et al.***

En base a estudios psicofísicos previos, se conoce que la percepción humana del contenido en frecuencia de una señal sigue una función no lineal, llamada la escala de *mel*, donde:

$$f_{mel} = 1125 \ln \left(1 + \frac{f}{700} \right)$$



Con lo anterior, es posible diseñar un banco filtros pasabanda linealmente espaciados en la representación de *mel* que, al filtrar una señal, se espera segmenten contenido de igual percepción en frecuencia.

Metodologías Actuales

El principal conjunto de características a obtener de la respuesta de la señal a cada filtro de frecuencias de mel , $e_i[m]$, son los coeficientes cepstrales:

$$C_i = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} \ln e_i[n+1] \cos \left[i \left(\frac{2n-1}{2} \right) \frac{\pi}{N} \right]$$

En forma general, el cepstrum corresponde a la transformada de Fourier discreta del logaritmo natural de una función en frecuencia. En este paper utilizan DCT en vez de FT.

En la señal resultante, los coeficientes C_i de mayor magnitud representarán la existencia de armónicos periódicos de magnitud f_s/C_i en el espectro original, donde f_s es la frecuencia de muestro de la señal (e.g. 44100Hz en audio).

Este vector de features se conoce como **Mel-Frequency Cepstral Coefficients (MFCC)**.

Metodologías Actuales

En el paper de Shih-Hao *et al.* se ocupan las características anteriores calculadas en **ventana dinámicas de Hamming**, las que permiten analizar la señal por tramos enfatizando las magnitudes del centro de la señal recortada.

Respecto de la clasificación, el conjunto de *features* extraídos es procesado por un clasificador SVM no lineal, pues se corroboró que un **Kernel exponencial de base radial (ERBF)** maximizaba la precisión del algoritmo.

Dado que el clasificador SVM esta diseñado para clasificar entre dos clases, se optó por un clasificador compuesto para análisis **uno-contra-uno** para cada par de clases, en vez de un formato uno-contra-todos.

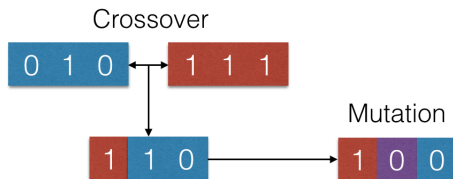
El conjunto de clases dadas por los distintos géneros musicales contenidos en el dataset utilizado fueron $\{classical, dance, lullaby, bossa, piano, blues\}$. Con todo lo anterior se logró una precisión del 98 % al entrenar con 45 canciones de un total de 300.

Metodologías Actuales

- **“Music Genre Classification Using GA-Induced Minimal Feature-Set”**, Nayak & Bhutani.

Es posible hallar un set mínimo y óptimo de características a utilizar en la clasificación de géneros musicales.

Para ello se utiliza un algoritmo genético, con recombinación y mutación aleatorias, que modela en cada bit de los cromosomas la utilización o no de una característica específica.



Metodologías Actuales

La selección de cromosomas en cada etapa se realizar por una función de *fitness* del tipo:

$$Fitness = HR + \gamma(CD/p) + \beta(c/p)$$

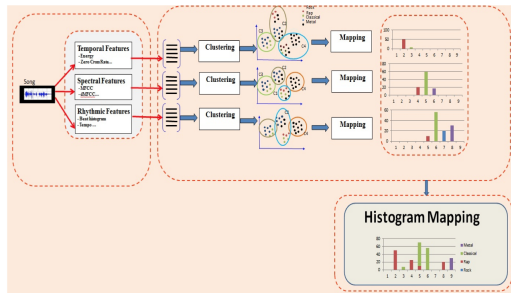
donde se tiene:

- *HR*: *Hit-rate*, es decir, la fracción de datos de entrenamiento correctamente clasificados.
- *CD*: distancia entre clases, para evitar mínimos locales por *HR*.
- *p*: número de muestras.
- *c*: número de ceros en el cromosoma, es decir, características no usadas.
- γ, β : factores de escalamiento, importancia de la separación de clases y de la minimización de características, respectivamente.

Metodologías Actuales

- “Feature Mapping and Fusion for Music Genre Classification”, Balti & Frigui.

Se presenta un concepto similar al *bag of words* donde las características se asocian en subconjuntos y, para cada uno, se obtiene una clusterización de las K clases.



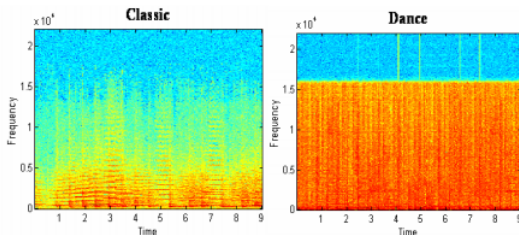
Luego, para cada cluster se genera una función de pertenencia gaussiana en torno al centro, la cual permite realizar un *soft assignment* con el cual generar un histograma normalizado, en base a datos de entrenamientos, que se utiliza como vector de características final.

Metodologías Actuales

- **“Music genre recognition based on visual features with dynamic ensemble of classifiers selection”**, Costa *et al.*

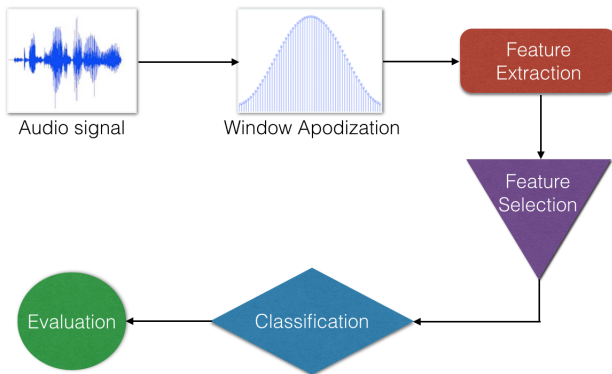
Las pistas de audio son convertidas a su correspondiente espectro en frecuencia, la imagen obtenida se utiliza como base de análisis. Se utilizan ventanas de audio en lugar de analizar el audio completo.

El entrenamiento se hace utilizando SVM entrenada con texturas utilizando LBP. El avance principal de este estudio se base en generar un conjunto dinámico de clasificadores que se ajusta a cada zona extraída del espectro.



Metodología a Investigar

La metodología que hemos decidido utilizar se compone por varias etapas presentadas en el siguiente diagrama.



Metodología a Investigar

Para cada una de las etapas se considera utilizar los siguientes algoritmos:

1. **Window Apodization:** ventana móvil de 10 segundos con estructura de Hamming para el énfasis en valores centrales.
2. **Feature Extraction:** Mel-Frequency Cepstral Coefficients, energía de la señal, frecuencia fundamental, Loudness Sensation, Integral Loudness, Zero Crossing Rate, entre otros.
3. **Feature Selection:** Principal Componente Analysis; preferimos una solución determinística antes de un algoritmo genético que pudiera inducir óptimos locales y sesgar los resultados.
4. **Classification:** Support Vector Machines, donde se probará la diferenciación de clases en formato uno-contra-uno y uno-contra-todos.
5. **Evaluation:** Se considerará tanto la precisión en la clasificación por ventanas como en una pista de audio completa. El primer caso es el típico de la literatura mientras que el segundo es el necesario de implementar en softwares de música como son iTunes y Spotify.

Datasets a Utilizar

- **GTZAN Genre Collection**

Este Dataset es utilizado en varios de los papers investigados. Consta de 1000 pistas de audio de 30 segundos cada una. La clasificación esta hecha sobre 10 géneros, por lo que hay 100 pistas por género. El formato de estas pistas es .au Los géneros utilizados son:

Rock, Reggae, Pop, Metal, Jazz, Hip-Hop, Disco, Country, Classical y Blues.

- **Music Audio Benchmark Dataset**

Este dataset consta de 1886 canciones en formato .mp3, donde nuevamente corresponden a una fracción de la pista de 10 segundos. La cantidad de pistas por género no estan definidas. Las clases utilizadas son las siguientes:

Alternative, Blues, Electronic, Folk-Country, Folk-Soul, Jazz, Pop, Rap Hip-Hop y Rock.

Bibliografía

- Shih-Hao Chen; Shi-Huang Chen; Guido, R.C., *"Music Genre Classification Algorithm Based on Dynamic Frame Analysis and Support Vector Machine"*, Multimedia (ISM), 2010 IEEE International Symposium on , vol., no., pp.357,361, 13-15 Dec. 2010
- Nayak, S.; Bhutani, A., *"Music Genre Classification Using GA-Induced Minimal Feature-Set"*, Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2011 Third National Conference on , vol., no., pp.33,36, 15-17 Dec. 2011
- Balti, H.; Frigui, H., *"Feature Mapping and Fusion for Music Genre Classification"*, Machine Learning and Applications (ICMLA), 2012 11th International Conference on , vol.1, no., pp.306,310, 12-15 Dec. 2012
- Costa, Y.; Oliveira, L.; Koerich, A.; Gouyon, F., *"Music genre recognition based on visual features with dynamic ensemble of classifiers selection"*, Systems, Signals and Image Processing (IWSSIP), 2013 20th International Conference on , vol., no., pp.55,58, 7-9 July 2013