Reconocimiento de Emociones basado en el contenido de la Música: Análisis de la letra de las canciones

Miguel Bustamante Cayo
School of Computer Science
Universidad Católica San Pablo
Perú - Arequipa
Email: miguel.bustamante@ucsp.edu.pe

***Resumen*—La música siempre ha estado íntimamente ligada a las emociones humanas, tal es así que en los últimos años ha surgido una tendencia, tanto comercial como de aplicación en la vida real, a agrupar las canciones en clases según las emociones que expresa su sonido y su letra. En el presente trabajo se desarrolla un análisis de las técnicas existentes (algoritmos) y los diferentes enfoques que se le da a la extracción de características.**

***Index Terms*—MIR, Music Information Retrieval, música, emociones, reconocimiento de emociones, análisis de emociones, análisis de sentimientos, análisis semántico.**

I. Introducción

La presencia de la música siempre ha sido una constante en la vida de las personas, sin embargo no es hasta la década pasada que gracias al rápido crecimiento de internet la música se ha vuelto uno de los elementos más importantes para los usuarios (desde usuarios comunes hasta empresas); hoy en día existen muchas aplicaciones basadas exclusivamente en la música, que van desde sitios web que ofrecen amplias listas de reproducción hasta redes sociales dedicadas exclusivamente a la música (como por ejemplo last.fm, pandora e incluso YouTube en menor grado).

Como resultado en la actualidad se cuenta con millones de canciones alojadas de forma digital en la nube [1], creando de esta manera la necesidad de obtener información de éstas que sean de interés para el usuario; de este manera una nueva disciplina dentro de Recuperación de Información (IR) toma importancia: Recuperación de Información de la Música (MIR), la cual usa la música como conjunto de datos [2], [3].

MIR involucra muchos campos de investigación (como por ejemplo musicología, psicología, análisis de señales, aprendizaje de máquina, etc.) y tiene muchas aplicaciones; dentro de los numerosos campos de investigación (como por ejemplo clasificación por género, sistemas de recomendación, transcripción, etc.) se encuentra el reconocimiento de emociones [4] [1], [3]

Dado que la música siempre ha estado ligada estrechamente a las emociones [4], el reconocimiento automático de emociones ha tomado un papel importante dentro de la Recuperación de Información de la Música; sin embargo no es una tarea sencilla y dado que aún no se ha logrado un nivel de precisión lo suficientemente alto, las empresas cuya idea de negocio gira entorno a este campo, prefieren etiquetar manualmente las canciones, pero se enfrentan a un problema: la constante aparición de nuevos temas musicales, por lo que deben actualizar constantemente sus bases de datos de manera manual (contratar especialistas que se encarguen de la clasificación) [5].

Dentro del reconocimiento automático de emociones se cuenta con sub-áreas de investigación que básicamente son las distintas formas mediante las cuales podemos reconocer emociones: la primera forma es mediante el análisis de las señales de audio, la segunda mediante el análisis de la letra de las canciones y el ultimo que utiliza datos externos (como comentarios de usuarios en internet) [4]

En este documento nos centraremos en el estudio del reconocimiento automático de emociones mediante el análisis del contenido de las letras canciones, para lo cual es necesario realizar un análisis del texto(letras de canciones) estableciendo relaciones y extrayendo caracteristicas para luego aplicar algún algoritmo de clasificación [6].

El presente estudio esta ordenado de la siguiente manera: en la primera sección se detallara la Recuperación de Información y la Recuperación de Información de la Música, luego se ahondara en el campo de análisis semántico, seguidamente se verá una de las partes mas importantes, la cual es la extraccion de caracteristicas. Por ultimo revisaremos el estado del arte en el reconocimiento automático de emociones en la letra de la musica, revisando los trabajos más relevantes y en los que se haya obtenido mejores resultados.

II. Recuperación de Información: Recuperación de Información de la Música

La Recuperación de Información se refiere a la recuperación de registros no estructurados, es decir datos que se encuentran

en forma libre, (un ejemplo de esto es el lenguaje natural), si bien la Recuperación de Información abarca muchos tipos de datos no estructurados como son imagenes, sonido, videos, etc. Su principal área de investigación es el trabajo con datos no estructurados del tipo textual(lenguaje natural). Un termino importane de conocer es el de documento, se llama documentos a los registros manejados por la Recuperación de Información [7].

La Recuperación de Informacion de la Musica es un campo de investigación especifico dentro de la recuperación de información, cuyo conjunto de datos no estructurado es la música. Este campo tiene muchas aplicaciones que estan basadas en el contenido, como son: Los sistemas de recomendación, segmentación, transcripción automática, clasificación por genero, reconocimiento automatico de emociones, etc [1], [3].

II-A. Áreas de estudio de Recuperación de Informacion de la Música

- Recuperación de Música
- Transcripción automática
- Segmentación
- Clasificación por género
- Sistemas de recomendación
- Detección de emociones

III. Detección de emociones

La música se considera como una forma de contenido ligado fuertemente a las emociones, ya que transmite ciertas emociones para el oyente, y por lo tanto se puede clasificar con etiquetas de emociones. Ser capaz de detectar automáticamente las emociones transmitidas por cualquier pista musical determinada es una tarea muy difícil que ha recibido considerable atención en la comunidad de recuperación de información de la música. Además, dado que se trata de la percepción humana de la música y del modelado de emociones, el tema del reconocimiento automático de emociones está estrechamente relacionado con el campo de la psicología de la música y de las personas. El reconocimiento automático de emociones en la música se puede aplicar tanto a la recuperación de la música y la recomendación. Por ejemplo, los usuarios pueden buscar música que transmite emociones particulares, o es posible que quieran recibir recomendaciones musicales en base a su estado de ánimo actual.

En [4] se hace una extensa revisión del estado del arte en reconocimiento de la emociones de la música. Cubre los temas de investigación de la psicología en las emociones, recogiendo anotaciones de emociones para el contenido de la música, el análisis de audio basado en el contenido y el análisis multimodal para el reconocimiento de emociones.
Existen muchas formas para reconocer emociones en una pista musical dentro de las principales tenemos:

- Reconocimiento de emociones a partir del audio.
- Reconocimiento de emociones a partir de la letra de la canción.
- Reconocimiento de emociones a partir de Datos proporcionados por medios externos.

Los tres enfoques mencionados suelen ser combinados para obtener una mayor precisión, en este estudio nos enfocaremos especialmente al reconocimiento de emociones basado en la letra de las canciones, para esto necesitamos aplicar conocimientos de análisis semántico, similitud semántica y algoritmos de clasificación y clusterización.

IV. Análisis Semántico

IV-A. Nociones generales y definicioness

En el campo del análisis semántico las definiciones de términos como similitud semántica, relación semántica y distancia taxonómica son variables entre distintos autores, por lo tanto tomaremos como base los conceptos presentados en [6], los cuales serán usados en subsecciones posteriores de la presente sección.

IV-A1. Medidas Semánticas: Los humanos durante nuestro proceso de aprendizaje somos capaces de relacionar cualquier cosa, relacionamos hechos, objetos y palabras, lo cual ayuda a construir nuevas experiencias y adquirir nuevos conocimientos. Entonces no es difícil darse cuenta que la relación y similitud son dos componentes centrales al momento de asignar algo a una categoría o reconocer patrones en las cosas.

Dicho lo anterior es fácil darse cuenta que las medidas semánticas son un aspecto muy importante, ya que pueden expresar cuan fuerte es la relación entre dos elementos (por ejemplo palabras y conceptos) basándose en el significado de cada uno. En las personas la percepción de la similitud está sujeta a diferentes factores como el conocimiento previo que tiene un individuo (por ejemplo un adolescente no tendrá la misma opinión que un anciano al momento de relacionar algo dentro de un ámbito tecnológico) o el contexto en que se encuentre.

En [6] se define las Medidas Semánticas como herramientas matemáticas usadas para estimar cualitativa y cuantitativamente el grado de relación entre dos unidades de lenguaje, conceptos o instancias, dando como resultado un valor numérico o simbólico. Dicho de otra manera lo que hacen las medidas semánticas es intentar imitar la capacidad ya mencionada de las personas para relacionar cosas, sin embargo esto es una tarea muy difícil, ya que es necesario simular un conocimiento previo en la máquina o en otros casos establecer criterios de relación.

Para el presente estudio se busca analizar letras de canciones, por lo tanto es necesario establecer una base de conocimiento para que la computadora sea capaz de clasificar de forma automática las canciones, para esto es necesario comprender correctamente dos nociones centrales dentro de las medidas semánticas: relación semántica y similitud semántica.

IV-A2. Relación Semántica y Similitud Semántica: Muchos autores consideran relación semántica y similitud semántica como dos conceptos iguales sin embargo podemos diferenciar ambos conceptos bajo un criterio muy simple: mientras que la relación semántica es algo general, la similitud semántica es la particularización de esta. Explicándolo más claramente, la relación semántica expresa el grado de relación entre dos elementos sin limitarse a un tipo de enlace semántico especifico (incluye relaciones léxicas y funcionales), por otro lado la similitud semántica especializa esto (solo incluye las relaciones léxicas), considerando solo la taxonomía u ontologías para dominios específicos [6], [8].

- **Taxonomía.-** Se define tradicionalmente como la ciencia de la clasificación pero en este caso lo que nos interesa es la similitud semántica en una taxonomía, la cual se define como la distancia entre dos elementos que están siendo comparados, en este caso es necesario definir una taxonomía basada en los conceptos de los elementos para luego encontrar la distancia taxonómica entre ambos elementos.
- **Ontología.-** Es un esquema conceptual dentro de uno o varios dominios, que en este caso define el vocabulario de un área mediante un conjunto de términos básicos y relaciones entre dichos términos, así como las reglas que combinan términos y relaciones que amplían las definiciones dadas en el vocabulario.

IV-B. Medidas de similitud semántica

A partir del surgimiento de métodos computacionales en las últimas décadas, se ha intentado obtener un método automatizado, fiable y eficiente para el cálculo de similitud semántica entre elementos. Las medidas resultantes pueden ser utilizadas para una gran variedad de aplicaciones dentro del campo del análisis semántico, que hoy en día están tomando gran importancia debido a las inmensas cantidades de información a la que tenemos acceso gracias a Internet. Con el objetivo de encontrar la medida de similitud que más se adapte al desarrollo del presente estudio en esta subsección se verá las principales ideas ya estudiadas y las medidas de similitud que se han presentado hasta el momento.

IV-B1. Medidas basadas en corpus: Aparecieron alrededor de los años 60 y son las medidas más antiguas, sin embargo no fue hasta los años 80 que se pudieron utilizar adecuadamente gracias al avance del hardware que hizo posible un procesamiento más eficiente de datos. Estas medidas están más enfocadas a medir relaciones semánticas en general que la similitud semántica.

Como primer conjunto de medidas tenemos las basadas en diccionarios o glosarios, aquí es donde Michael Lesk en [9] plantea un algoritmo para obtener el significado de una palabra dentro de un contexto específico, su idea básicamente consiste en que si un par de palabras tienen relación compartirán palabras en su definición. Es decir para desambiguar una palabra en una oración se escoge la definición que comparta más palabras con el resto de palabras de la oración.

Estas medidas basadas en corpus tienen un problema, el cual es que están muy orientadas a medir la relación semántica mas no la similitud, además que para obtener un grado de precisión lo suficientemente confiable es necesario acotar el dominio a algo muy específico y utilizar el corpus adecuado para dicho dominio, con un dominio muy amplio o con el corpus inadecuado los resultados no son muy aceptables.

IV-B2. Medidas basadas en grafos: Al contrario que las medidas basadas en corpus, en las medidas basadas en grafos se utiliza estructuras jerárquicas (taxonomías), en la cual un nodo representa a un concepto y los nodos están unidos por diferentes tipos de enlaces que representan los diferentes tipos de relaciones. Una de las relacionas más usadas es la de hiperonimia-hiponimia (establece la relación “es un”, donde el nodo hijo es un subgrupo del nodo padre), pero también existen otras relaciones como la de meronimia-holonimia (expresa la relación “parte de”) o la de sinonimia (expresa la relación de igualdad de significado).

Una de las taxonomías más usadas en el campo del análisis semántico es Wordnet, Wordnet es la principal referencia en cuanto a herramientas para la identificación taxonómica y las relaciones entre conceptos gracias a su creciente alcance y a que es de licencia libre (Wordnet solo está disponible para el idioma ingles sin embargo tiene una adaptación llamada EuroWordnet que cuenta con soporte para distintos idiomas, entre ellos el español).

IV-B3. Medidas basadas en múltiples fuentes de información: Estas medidas son una especie de combinación de las anteriormente detalladas, mezclan la utilización de un corpus y una jerarquía para obtener mejores resultados. Su principal eje es aumentar la información contenida en una taxonomía (jerarquía de conceptos) para obtener una mayor precisión al momento de hallar la similitud semántica.

Se han propuesto diversas técnicas que utilizan una jerarquía, en su mayoría Wordnet, junto con un corpus (por ejemplo Brown Corpus para el ingles), sin embargo siguen arrastrando el problema de las medidas basadas en corpus, ya que necesitan de un corpus específico para el dominio si se quieren obtener buenos resultados.

IV-B4. Medidas basadas en buscadores web: Como consecuencia del gran avance de los buscadores web (por ejemplo Google, Yahoo, Bing, Ask, etc.) tanto en eficiencia como en eficacia, ha surgido una nueva opción para encontrar medidas de similitud semántica, básicamente su utilización emula de alguno u otra manera una especie de corpus, dado

su gigantesco alcance y la inmensa cantidad de información es posible tratar palabras que no se pueden encontrar en los corpus tradicionales, siendo esta su mayor ventaja.

Los trabajos en esta área son un poco recientes tal es así que en el 2006 Strube y Ponzetto [10] realizan una medida muy básica que consiste en tomar los hits (número de resultados) luego de realizar la búsqueda en un buscador y aplicarles la formula de la *ecuación 1*.

$$Sim(W1, W2) = \frac{Hits(w1, w2)}{Hits(w1) + Hits(w2) - Hits(w1, w2)} \quad (1)$$

Por otro lado existen algunos inconvenientes con este tipo de medidas, como el hecho que miden relaciones en general y no la similitud semántica, sumado a esto el hecho de que dos palabras pueden aparecer en un mismo documento pero no necesariamente tendrían algo que ver, además que aquí se ve bien evidenciado el problema de la polisemia ya que el conteo de páginas incluye los diversos significados que esta pueda tener. Este conjunto de medidas basadas en buscadores web no han tenido mayor prosperidad ya que su nivel de precisión está por debajo que las antes vistas.

IV-B5. Medidas basadas en Wikipedia: Un punto medio entre los buscadores (enormes cantidades de información desestructurada) y jerarquías como Wordnet lo podemos encontrar en Wikipedia. Wikipedia tiene una gran cantidad de entradas clasificadas en diversas taxonomías representando una especie de herencia múltiple, además contiene ciclos dentro de su estructura.

Es claro que Wikipedia no ha sido diseñada con una estructura jerárquica rigurosa sin embargo es posible tratarla como tal para encontrar relaciones de similitud, cabe destacar que si bien la mayoría de jerarquías y corpus solo están disponibles para el inglés, Wikipedia cuenta con versiones para muchos lenguajes y que la cantidad de información allí guardada crece constantemente gracias al apoyo de la comunidad.

V. Modelos para la Clasificación de Emociones

En la clasificación de emociones es necesario el uso de un modelo psicológico de emociones, por esto algunos investigadores definieron modelos de acuerdo a una o más dimensiones. A continuación se verá algunos de los principales modelos, algunos de los cuales son utilizados como base para la clasificación de emociones en trabajos relacionados.

V-A. Modelo Circumplejo

El modelo circumplejo de emoción fue desarrollado por James Russell. Este modelo sugiere que las emociones están distribuidas en un espacio circular de dos dimensiones, que contiene dimensiones de excitación y de valencia. La excitación representa el eje vertical y la valencia representa el eje horizontal, mientras que el centro del círculo representa una valencia neutra y un nivel medio de la excitación. En este modelo, los estados emocionales pueden ser representados en cualquier nivel de valencia y excitación, o en un nivel neutro de uno o ambos de estos factores [11].

V-B. Modelo Vectorial

El modelo vectorial apareció por primera vez en 1992. Éste es un modelo bidimensional formado por vectores que apuntan en dos direcciones, lo que representa una forma de ”boomerang”. El modelo asume que siempre hay subyacente una dimensión excitación, y que la valencia determina la dirección en la que se encuentra una emoción particular. Por ejemplo, una valencia positiva desplazaría la emoción hasta el vector de la parte superior y una valencia negativa desplazaría la emoción hacia abajo. En este modelo, los estados de excitación altos se diferencian por su valencia, mientras que los estados de excitación bajos son más neutrales y están representados cerca del punto de encuentro de los vectores [11].

V-C. Modelo Activación Positiva - Activación Negativa (PANA)

El modelo de activación positiva - activación negativa (PANA) o modelo çonsensual”de la emoción, originalmente creado por Watson y Tellegan en 1985, sugiere que el afecto positivo y el afecto negativo son dos sistemas separados. Al igual que en el modelo vectorial, los estados de mayor excitación tienden a ser definido por su valencia, y los estados de menor excitación tienden a ser más neutros en términos de valencia. En el modelo PANA, el eje vertical representa menor a mayor afecto positivo y el eje horizontal representa bajo a alto afecto negativo. Las dimensiones de valencia y excitación se ponen en una rotación de 45 grados sobre estos ejes [11].

V-D. Modelo de Plutchik

Robert Plutchik ofrece un modelo tridimensional que es un híbrido, arregla las emociones en círculos concéntricos donde los círculos interiores son más básicos y círculos exteriores más complejos. En particular, los círculos exteriores también se forman mediante la mezcla de los círculos internos. El modelo de Plutchik, según Russell, emana de una representación circumpleja, donde se representaron las palabras emocionales basadas en la similitud. Este modelo es muy usado en computación para tareas como la interacción humana-afectiva con el ordenador o análisis de sentimientos [11].

VI. Extracción de Características

Una de las partes más importante en el reconocimiento automático de emociones es la extracción de características representativas de letra de cada canción por tal motivo en

esta sección revisaremos las formas más comunes de extraer características a partir de las letras de las canciones.

VI-A. Pre procesamiento de texto

Según sea el enfoque que se le esté dando a la extracción de características en los documentos se puede usar o no usar las siguientes técnicas para el pre procesamiento del texto (en este caso la letra de las canciones).

Tokenización.- Es el proceso de dividir un flujo de texto en palabras, frases, símbolos u otros elementos significativos llamados tokens. La obtención de dichos tokens tiene como objetivo preparar el texto para una tarea posterior, tales como el análisis o la minería de texto. La tokenización es útil tanto en la lingüística (donde es una forma de segmentación de texto) y en ciencias de la computación, donde forma parte del análisis léxico (muy fuertemente ligada a la minería de texto). Limpieza de signos de puntuación.- Consiste en remover todos los signos de puntuación del texto, ya que en la mayoría de casos es irrelevante para las operaciones de minería de texto. Limpieza de Stop Words.- Consiste en remover palabras que se repiten con mucha frecuencia y que no son relevantes para la extracción de características(a dichas palabras se les denomina stop words), estas palabras son artículos, conjunciones, adverbios, etc.

Lematización.- La lematización es un proceso lingüístico que consiste en, dada una forma flexionada (es decir, en plural, en femenino, conjugada, etc), hallar el lema correspondiente. El lema es la forma que por convenio se acepta como representante de todas las formas flexionadas de una misma palabra. Es decir, el lema de una palabra es la palabra que nos encontraríamos como entrada en un diccionario tradicional: singular para sustantivos, masculino singular para adjetivos, infinitivo para verbos. Hay diversos grados de lematización posible: podemos hacer una lematización puramente morfológica, o bien hacer una lematización sintáctica que tenga en cuenta el contexto en el que aparece la palabra. La lematización es utilizada en buscadores, traductores automáticos, extracción de información y demás herramientas vinculadas al Procesamiento del Lenguaje Natural.

Stemming.- Este método es muy similar a la lematización sin embargo la diferencia radica en que este es un método para reducir una palabra a su raíz (en inglés stem) a diferencia de la lematización que la reduce a su lema. Hay algunos algoritmos de stemming que ayudan en sistemas de recuperación de información. Stemming aumenta el recall que es una medida sobre el número de documentos que se pueden encontrar con una consulta[XU].

VI-B. Obtención de características a partir de Frameworks existentes

jLyrics.- Es un conjunto de herramientas de software para la minería automática de letras de canciones en internet y para la extracción de características a partir de las letras, una vez que estas se han adquirido. Permite extraer características como: número de palabras, frecuencias, número de líneas, frecuencia de la puntuación, bigramas, etc.

Synesketch.- Analiza el contenido emocional de las frases de texto en términos de tipos de emociones (alegría, tristeza, ira, miedo, disgusto y sorpresa), pesos (la intensidad de la emoción), y un valor (¿es positivo o negativo). La técnica de reconocimiento se basa en un método refinado de identificación de palabras clave que emplea un conjunto de heurísticas, una léxico de palabras basado en WordNet, y un léxico de los emoticonos y abreviaturas comunes.

ConceptNet.- Se expresa como un grafo dirigido cuyos nodos son conceptos, y cuyas aristas son afirmaciones de sentido común acerca de estos conceptos. Los conceptos representan conjuntos de frases en lenguaje natural más cercanos, que podrían ser sintagmas nominales, sintagmas verbales, frases adjetivas o cláusulas.

VI-C. Bag of Words

El modelo de Bag of Words (bolsa de palabras) es una representación simplificada utilizada en el procesamiento de lenguaje natural y en la recuperación de información (IR). En este modelo, un texto (por ejemplo, una frase o un documento) se representa como una bolsa (conjunto múltiple) de palabras, sin tener en cuenta la gramática e incluso orden de las palabras, pero manteniendo la multiplicidad. El modelo de bolsa de palabras se utiliza comúnmente en los métodos de clasificación de documentos, donde se utiliza la ocurrencia de cada palabra (frecuencias) como una característica para el entrenamiento de un clasificador. [12]

A partir de la bolsa de palabras se forma vectores que representen los documentos, dichos vectores pueden ser construidos a partir de valores booleanos (solo indican si la palabra se encuentra en el documento) o valores de pesos (frecuencias). Los tipos de frecuencias más usadas ser el "term frequency" y el "Term frequency – Inverse document frequency". [13]

TF.- Es una medida que expresa la frecuencia de un término en un documento, esta puede ser en general, es decir simplemente el número de veces que aparece el termino en el documento, o también puede ser normalizada lo cual estaría dado por su frecuencia dividida entre la frecuencia máxima tal y como se observa en la *ecuación 2*, donde t es el termino y d es el documento. [13]

$$tf(t,d) = \frac{f(t,d)}{max\{f(w,d) : w \in d\}} \quad (2)$$

TF-IDF.- Es una medida numérica que expresa cuán relevante es una palabra para un documento en una colección. Esta medida se utiliza a menudo como un factor de ponderación en la recuperación de información y la minería de texto; este valor aumenta proporcionalmente al número de veces que una palabra aparece en el documento, pero es compensada por la frecuencia de la palabra en la colección de documentos, lo que permite manejar el hecho de que algunas palabras son generalmente más comunes que otras. Dada la *ecuación 2* donde se observa el cálculo del tf, y la *ecuación 3*. donde se observa el cálculo del idf (número total de documentos entre el número de documentos en los cuales aparece el término), obtenemos la *ecuación 4* donde el tf-idf es el producto del tf por el idf. [13]

$$idf(t,D) = log\frac{/D/}{/\{d \in D : t \in d\}/} \quad (3)$$

$$tfidf(t,d,D) = tf(t,d) \times idf(t,D) \quad (4)$$

VII. Técnicas utilizadas en el Reconocimiento Automático de Emociones

VII-A. Métodos no Supervisados

VII-A1. ***K-Means:*** Es un algoritmo de agrupamiento(clusterización), que tiene como objetivo la partición de n documentos en k grupos en el que cada documento pertenece al grupo más cercano a la media, este es uno de los métodos más utilizados en minería de datos.

El problema es computacionalmente difícil (NP-hard). Sin embargo, hay heurísticas eficientes que se emplean comúnmente y convergen rápidamente.

Los pasos del algoritmo k-means son los siguientes:

1. Escoger aleatoriamente nuestros centroides (el mismo número de centroides que el número de grupos que queremos hallar). K centroides iniciales.

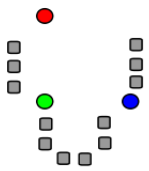

Figura 1. Se escogen aleatoriamente los centroides

2. K grupos son generados asociando el centroide más cercano a cada documento.

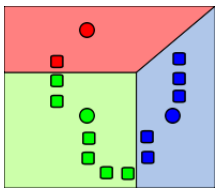

Figura 2. Grupos generados

3. El centroide de cada grupo se recalcula.

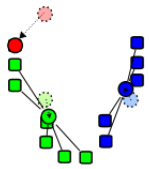

Figura 3. El centroide se recalcula

4. Los dos pasos anteriores se repiten hasta lograr la convergencia.

VII-B. Métodos Supervisados

VII-B1. ***K Nearest Neighbors:*** Este es un método de clasificación no paramétrico, que estima el valor de la función de densidad de probabilidad o directamente la probabilidad a posteriori de que un elemento "x" pertenezca a la clase "C" a partir de la información proporcionada por el conjunto de prototipos. La fase de entrenamiento del algoritmo consiste en almacenar los vectores característicos y las etiquetas de las clases de los elementos del conjunto de entrenamiento. En la fase de clasificación, la evaluación del elemento a clasificar es representada por un vector en el espacio característico. Se calcula la distancia entre los vectores almacenados y el nuevo vector, y se seleccionan los k ejemplos más cercanos, de esta manera el nuevo elemento es clasificado con la clase que más se repite en los vectores seleccionados. Este método supone que los vecinos más cercanos nos dan la mejor clasificación y esto se hace utilizando todos los atributos; el problema de dicha suposición es que es posible que se tengan muchos atributos irrelevantes que dominen sobre la clasificación: dos atributos relevantes perderían peso entre otros veinte irrelevantes.

VII-B2. ***Naive Bayes:*** En español conocido como "Bayes Ingenuo", es un clasificador probabilístico fundamentado en el teorema de Bayes y algunas hipótesis simplificadoras adicionales. Es a causa de estas simplificaciones, que se suelen resumir en la hipótesis de independencia entre las variables predictoras, que recibe el apelativo de ingenuo. Un clasificador de Bayes ingenuo asume que la presencia o ausencia de una característica particular, no está relacionada con la presencia o ausencia de cualquier otra característica, dada la clase variable. El clasificador de Bayes ingenuo esta dado por la formula de probabilidad 5 y la formula 6.

$$P(a_1, .., a_n) = \prod_i^n P(a_i/v_j) \quad (5)$$

$$\boldsymbol{v}_{NB} = argmax P(\boldsymbol{v}_j) \prod_i^n P(a_i/v_j) \quad (6)$$

VII-B3. ***Support Vector Machine****:* Es un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik [14] y su equipo en los laboratorios AT&T. Una SVM es un modelo que representa a los puntos de muestra en el espacio, separando las clases por un espacio lo más amplio posible.

Dado un conjunto de puntos, subconjunto de un conjunto mayor (espacio), en el que cada uno de ellos pertenece a una de dos posibles categorías, un algoritmo basado en SVM construye un modelo capaz de predecir si un punto nuevo (cuya categoría desconocemos) pertenece a una categoría o a la otra. Como en la mayoría de los métodos de clasificación supervisada, los datos de entrada (los puntos) son vistos como un vector p-dimensional.
La SVM busca un hiperplano que separe de forma óptima a los puntos de una clase de la de otra, que eventualmente han podido ser previamente proyectados a un espacio de dimensionalidad superior. En ese concepto de "separación óptima.[es] donde reside la característica fundamental de las SVM: este tipo de algoritmos buscan el hiperplano que tenga la máxima distancia (margen) con los puntos que estén más cerca de él mismo.

En la *Figura 4* se puede observar el hyperplano con distancia optima en el espacio de elementos, ademas de los vectores de soporte que son los utilizados para hallar la maxima distancia.

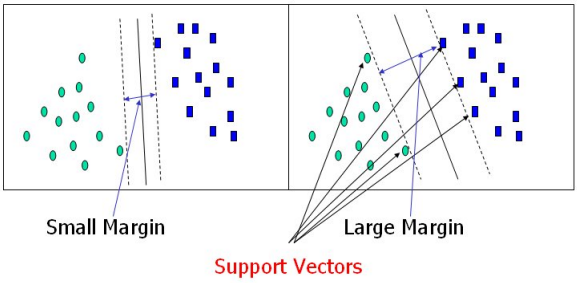

Figura 4. Support Vector Machine

VII-B4. ***Adaboost****:* Es la abreviación de "Adaptative Boosting", fue formulado por Yoav Freund y Robert Schapire; la idea básica de este algoritmo es combinar clasificadores débiles (baja presición) sumando las distintas salidas de los clasificadores débiles de una manera ponderada. Si bien los clasificadores débiles son de baja precisión, su rendimiento debería estar por encima de una clasificación aleatoria (es decir debería estar por encima del 50 % en un escenario con dos clases), de esta manera se asegura que el clasificador final sea fuerte. A diferencia de las redes neuronales y SVMs, el proceso de entrenamiento de AdaBoost selecciona sólo aquellas características que se sabe que mejoraran el modelo predictivo para mejorar el poder predictivo del modelo, de esta forma se reduce la dimensionalidad y se mejora el tiempo de ejecución. [15]

VIII. Trabajos Relacionados

La mayor parte de los estudios en la clasificación basado en emociones se basan en datos recogidos de fuentes en Internet. Estos conjuntos de datos se suelen pre-clasificar en emociones, a través de las etiquetas, tomadas por ejemplo de sitios como Last.FM o AllMusic [16]. En el caso de la extracción de características de las letras, las características más utilizadas son las características estadísticas como Bag of Words (BOW) [17]. Otros estudios incluyen la utilización de características estilísticas y lingüísticas del texto [18]. En esos estudios, las características son representadas por una serie de medidas, como por ejemplo, tf-idf o representación booleana [19]. En la mayoría de los estudios, las características de audio superan características líricas y la combinación de ambos suele proporcionar mejores resultados [16].

Dan Yang en [5] toma como base las 168 categorías emocionales de allmusic.com y a partir del modelo Activación Positiva – Activación Negativa hace una reducción a 23 emociones, sin embargo deja de lado algunas que no pueden ser reducidas, en la *Figura 5* se puede observar la reducción realizada. Para la extracción de características es usado "General Inquirer" (léxico psicolingüístico que contiene 182 categorías psicologías y más de 10000 palabras), de esta manera ya no se tiene un vector de características de dimensionalidad muy alta (por ejemplo 5000) ya que todas las palabras únicas son reducidas a 182 categorías.

Positive emotions		Negative emotions	
Allmusic	EMO	Allmusic	EMO
Brash, Bravado, Swaggering	Proud	Acerbic, Aggressive, Bitter, Fiery, Nasty, Outraged, Rebellious, Snide	Anger
Bright, Effervescent, Gleeful, Humorous, Irreverent, Party/Celebratory, Rambunctious, Raucous, Rollicking, Silly, Sweet, Whimsical, Witty	Joy	Aggressive	Aggressive
Yearning, Sexy, Provocative, Sleazy, Sensual	Arousal	Bleak, Distraught, Somber, Poignant, Melancholy, Plaintive	Sad
Exuberant	Exuberant	Gloomy	Gloomy
Passionate	Passionate	Cynical, Ironic	Alienation
Happy	Happy	Manic, Paranoid, Spooky, Unsettling	Fear
Wistful	Reflective	Ominous	Ominous
Romantic, Precious, Reverent, Sentimental, Warm	Love	Eerie	Eerie
Soothing	Soothing	Tense	Tense
Laidback/Mellow, Gentle, Refined/Mannered, Reserved, Restrained	Calm	Greasy	Disgust
Confident	Confident	Eccentric, Fractured, Trippy, Freakish, Spacey	Psychotic
Detached	Reflective	Aggressive, Boisterous, Brittle, Brooding, Cold, Confrontational, Harsh, Wry, Dramatic, Theatrical, Volatile, Thuggish, Trashy, Visceral, Earthy, Rustic	
Complex, Enigmatic, Eccentric, Street-smart, Stylish, Earnest, Ramshackle, Smooth, Slick, Sophisticated, Gritty, Spiritual, Searching, Playful, Rousing, Free-wheeling			

Figura 5. Reducción de emociones de allmusic.com a 23 tomando como base el modelo PANA

Rada Mihalcea en [20] hace un reconocimiento de emociones línea por línea para reconocer todas las emociones que puede contener una canción, se basa en las 6 emociones propuestas por P. Ekman en [21] para esto hace uso de dos enfoques al momento de obtener los resultados. El primero es el ya ampliamente conocido uso del bag of words haciendo uso de unigramas (según se explica el uso de bigramas, trigramas, etc. no mejora los resultados), y por otro lado se hace uso de clases semánticas (características lingüísticas)

al igual que en [5], solo que no se hace uso de "General inquirer", en su lugar se usan Linguistic Inquire and Word Count (LIWC) y Wordnet Affect (WA). En la *Figura 6* se pueden observar los resultados al aplicar los dos enfoques por separado o aplicarlos juntos.

Emotion	Unigrams	Semantic Classes	All Textual
ANGER	0.5525	0.3044	0.5658
DISGUST	0.4246	0.2394	0.4322
FEAR	0.3744	0.2443	0.4041
JOY	0.5636	0.3659	0.5769
SADNESS	0.5291	0.3006	0.5418
SURPRISE	0.3214	0.2153	0.3392
AVERAGE	0.4609	0.2783	0.4766

Figura 6. Resultados de los experimentos de Mihalcea

Seungwon propone utilizar solo la parte de introducción que crea una atmósfera específica de una canción, y la parte de coro que es la parte más fuerte de la canción. Se basa en el modelo de Plutchik [11] y en 8 emociones. Primero se coge la parte introductoria que tendrá las palabras más importantes y luego que el coro que repetirá constantemente las palabras esenciales, a esto se le aplica un conteo de palabras con Affective Norms for English Words (ANEW) que es otro lexico como en [5] y [20],además la similaridad entre las palabras y la emoción (similaridad semántica). [22]

Por ultimo en [23] se hace uso de Adaboost como alternativ al ya tradicional SVM, obteniendo mejores resultados que el ya mencionado.

IX. Pruebas

IX-A. Corpus

Para realizar las pruebas se construyó un corpus de letras de canciones, para esto se hizo un crawler que consistía de dos partes. La primera parte consiste en extraer canciones (título, autor, género, álbum, etc.) de AllMusic a través del API de rovicorp, estas canciones fueron extraídas por emociones (un total de 289), y aproximadamente se trajo 1000 canciones por emoción; la segunda parte consiste en extraer las letras de las canciones recuperadas de Allmusic, esto se hizo con el API proporcionado por Chartlyrics.com, sin embargo solo se logró recuperar alrededor de 8000 canciones en los primeros 25 grupos de canciones ordenados según su emoción.

IX-B. Evaluación

El corpus estuvo conformado por 22 emociones (las que se alcanzaron a crawlear), sin embargo se pudo observar que muchas canciones (2000 de las casi 8000) estaban en más de un grupo de emociones, por lo que en el futuro será necesario reordenar estas emociones. Para realizar las pruebas se usó SVM (con kernel lineal y rbf) y Adaboost con árboles de decisión, ambos multiclase, para la evaluación se utilizó validación cruzada. En general se obtuvieron muy malos resultados, cuyo origen probablemente está en el corpus de datos. SVM dio alrededor de 40 % de precisión, mientras que Adaboost alrededor de 45 %.

X. Conclusiones

Hasta el momento las técnicas que han dado mejores resultados son SVM y AdaBoost.

La extracción de caracteristicas es una de las partes mas importantes para la clasificación.

Referencias

[1] Michael A. Casey, Remco Veltkamp, Masataka Goto, Christophe Rhodes Marc Leman, and Malcolm Slaney. Content-based music information retrieval: Current directions and future challenges. *Proc. IEEE*, 96(4):668–696, 2008.

[2] Ricardo Jorge Sebastiao dos Santos. Music Information Retrieval: Developing Tools for Musical Content Segmentation and Comparison. Master's thesis, Universidade Técnica de Lisboa, Lisboa, Octubre 2010.

[3] Marius Kaminskas and Francesco Ricci. Contextual music information retrieval and recommendation: State of the art and challenges, Mayo 2012.

[4] Youngmoo E. Kim, Erik M. Schmidt, Raymond Migneco, On G. Morton, Patrick Richardson, Jeffrey Scott, Jacquelin A. Speck, and Douglas Turnbull. Emotion recognition: a state of the art review. In *11th International Society for Music Information and Retrieval Conference*, 2010.

[5] Dan Yang and Won-Sook Lee. Music emotion identification from lyrics. In *Multimedia, 2009. ISM '09. 11th IEEE International Symposium on*, pages 624–629, Dec 2009.

[6] Sébastien Harispe, Sylvie Ranwez, Stefan Janaqi, and Jacky Montmain. Semantic measures for the comparison of units of language, concepts or entities from text and knowledge base analysis. *CoRR*, abs/1310.1285, 2013.

[7] Ed Greengrass. Information retrieval: A survey, 2000.

[8] Ted Pedersen, Serguei V. S. Pakhomov, Siddharth Patwardhan, and Christopher G. Chute. Measures of semantic similarity and relatedness in the biomedical domain. *J. of Biomedical Informatics*, 40(3):288–299, June 2007.

[9] Michael Lesk. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proceedings of the 5th Annual International Conference on Systems Documentation*, SIGDOC '86, pages 24–26, New York, NY, USA, 1986. ACM.

[10] Michael Strube and Simone Paolo Ponzetto. Wikirelate! computing semantic relatedness using wikipedia. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2*, AAAI'06, pages 1419–1424. AAAI Press, 2006.

[11] David C. Rubin and Jennifer M. Talarico. A comparison of dimensional models of emotion: Evidence from emotions, prototypical events, autobiographical memories, and words. *Memory*, 17(8):802–808, 2009. PMID: 19691001.

[12] Yin Zhang, Rong Jin, and Zhi-Hua Zhou. Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1-4):43–52, 2010.

[13] Juan Ramos. Using tf-idf to determine word relevance in document queries. In *Proceedings of the First Instructional Conference on Machine Learning*, 2003.

[14] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[15] Yoav Freund, Raj Iyer, Robert E. Schapire, and Yoram Singer. An efficient boosting algorithm for combining preferences. *J. Mach. Learn. Res.*, 4:933–969, December 2003.

[16] Xiao Hu and J Stephen Downie. When lyrics outperform audio for music mood classification: A feature analysis. In *ISMIR*, pages 619–624, 2010.

[17] Fabrizio Sebastiani. Machine learning in automated text categorization. *ACM Comput. Surv.*, 34(1):1–47, March 2002.

[18] Xiao Hu and J Stephen Downie. Improving mood classification in music digital libraries by combining lyrics and audio. In *Proceedings of the 10th annual joint conference on Digital libraries*, pages 159–168. ACM, 2010.

[19] Xiao Hu. *Improving music mood classification using lyrics, audio and social tags*. PhD thesis, University of Illinois at Urbana-Champaign.

[20] Rada Mihalcea and Carlo Strapparava. Lyrics, music, and emotions. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, EMNLP-CoNLL '12, pages 590–599, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.

[21] Paul Ekman. Facial expression and emotion. *American psychologist*, 48(4):384, 1993.

[22] Seungwon Oh, Minsoo Hahn, and Jinsul Kim. Music mood classification using intro and refrain parts of lyrics. In *Information Science and Applications (ICISA), 2013 International Conference on*, pages 1–3, June 2013.

[23] Dan Su and P. Fung. These words are music to my ears: Recognizing music emotion from lyrics using adaboost. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific*, pages 1–4, Oct 2013.