

Rapport d'avancement de LIRIS

juin 2006

Nous avons proposé un classificateur de titres de musique en genre s'appuyant sur une architecture de comité d'experts où chaque expert utilise des caractéristiques spécifiques. Dans cette architecture des experts de classification par analyse acoustique, des experts de classification par analyse rythmique et des experts de classification par analyse textuelle coopèrent afin de fournir la classification globale de titres de musique. Chacun de ces experts produit une probabilité d'appartenance d'un titre de musique aux différentes classes, qui sont ici des genres musicaux. Chaque classificateur a un poids qui sera attribué par rapport à ses performances sur une base de développement. La somme des probabilités multipliées par les poids des experts constitue la sortie du classificateur général.

L'expert de classification par analyse acoustique est un ou plusieurs perceptrons multi-couches (MLP) ayant comme entrée les caractéristiques acoustiques issues d'une Modélisation Gaussienne Incrémentale (MGI) perceptuellement motivées. Chaque MLP de cet expert sera entraîné indépendamment des autres sur toute ou une partie des données d'apprentissage. La modélisation MGI s'appuie sur un modèle de mémoire auditive à court terme et est une modélisation du spectre du signal sonore inspirée par des phénomènes psychoacoustiques associée à une modélisation du filtre auditif humain.

L'expert de classification par analyse rythmique est un classificateur de type K-plus proches voisins (K-PPV) qui utilise le *beat histogram* à deux dimensions et une mesure de similarité basique. La probabilité d'appartenance d'un titre musical à une classe est proportionnelle au nombre de titres de la même classe compris dans les 15 plus proches voisins. Les caractéristiques rythmiques sont des caractéristiques obtenues à partir d'une détection de *beats*, ou *onsets*, d'un signal de musique. Nous avons développé une description rythmique que nous nommons *beat histogram* à deux dimensions. En effet, une détection des *beats* dans un signal de musique est réalisée en produisant une courbe de probabilités qu'il suffit de comparer à un seuil pour détecter les moments *beats* et ensuite en calculer la période. Un *beat histogram* est un histogramme des valeurs de période de *beats* d'un signal de musique. Il contient une information rythmique plus riche que le *bpm* (beats per minute). Un *beat histogram* à deux dimensions est la combinaison de plusieurs *beat histogram* pour plusieurs valeurs de seuil de détection des *beats*. Notre approche de détection de *beats* s'appuie sur une transformation en ondelettes continue (CWT) comme la représentation spectrale du signal. La CWT permet d'ajuster l'échelle temps/fréquence pour être plus précis au niveau de résolutions temporelle en hautes fréquences.

L'expert de classification par analyse textuelle est un classificateur qui s'appuie sur un crawler Web afin de classifier les artistes en genre. Le crawler envoie une requête qui contient le nom de l'artiste et le nom d'un genre de musique à un moteur de recherche (le moteur Google a été utilisé dans cette étude). Le nombre de documents retournés pour un couple nom d'artiste-nom d'un genre constitue une information sur la probabilité que l'artiste appartienne au genre en question. Nous pouvons ainsi

obtenir pour chaque artiste la probabilité d'appartenance aux genres de musique que nous définissons.

La combinaison des 3 experts, acoustique, rythmique et textuel, a apporté un gain de performance de classification considérable. Le taux de précision a passé de 52% pour le meilleur expert individuel à 63% pour la combinaison des deux meilleurs et à 70% pour la combinaison des trois. Nous pouvons conclure que le problème de classification de la musique en genre est un problème nécessitant la collaboration de plusieurs types de classificateurs et de caractéristiques. L'analyse acoustique trouve rapidement ses limites et elle doit être complétée par une analyse des caractéristiques d'ordre culturel.

Par ailleurs, la classification en genre de la musique a été aussi l'objet des travaux en commun entre LIRIS ECL et ENST. En effet, des premières expérimentations sur un système de classification commun utilisant des descripteurs développés dans chacune de nos deux équipes semblent indiquer que des performances ont pu être améliorées. Les résultats expérimentaux une fois stabilisés feront l'objet d'une communication du projet.

Parallèlement à cette étude, une autre approche a été développée dans le but de réaliser la classification de titres musicaux en genres. Les caractéristiques extraites du signal audio reposent sur son codage selon une analyse temporelle, fréquentielle et temps-fréquence. Ces codages permettent alors de représenter le signal comme une séquence de motifs dont la distribution est analysée selon les lois de Zipf. Les caractéristiques extraites par application de ces lois représentent les entrées du classifieur qui est constitué d'un ensemble de MLP organisés de manière hiérarchique de façon à respecter la taxonomie des classes.

Un autre aspect de notre travail concerne la définition de mesures de similarités musicales.

Une première mesure est une similarité rythmique correspondant à une distance entre deux *beat histograms* obtenus de manière identique à celle réalisée dans le cas de l'expert de classification par analyse rythmique.

Deux autres mesures ont été développées, représentant une similarité mélodique et reposant sur la détection de notes. Celle-ci utilise une technique inspirée de modèles harmoniques de périodes fondamentales (pitch) et se décompose en deux étapes. La première étape consiste à générer un modèle harmonique représentant un ensemble de pics successifs correspondant à la position de la fréquence fondamentale et de ses harmoniques, obtenu à partir d'un spectrogramme issu d'une analyse en ondelettes. La deuxième étape consiste en l'analyse d'un signal sonore. Ce dernier est traité par une analyse en ondelettes de manière à en extraire la structure harmonique, puis la corrélation entre cette structure et le modèle est évaluée. Une corrélation importante permettra d'en déduire la note correspondant à la fréquence fondamentale considérée. A partir de cette transcription, deux mesures de similarité mélodique ont été définies. La première repose sur les profils de notes (histogrammes) calculés sur l'ensemble du titre musical permettant d'en déduire la tonalité à partir de la distribution des notes. Deux titres musicaux sont alors comparés en calculant la similarité de leur profil de notes. La deuxième mesure de similarité musicale est basée sur les histogrammes de

succession de notes. Ainsi, le nombre d'occurrences des combinaisons de chaînes de trois notes est calculé de manière à en obtenir l'histogramme. Cet histogramme possède trois dimensions, chacune représentant une note de la chaîne. Deux titres musicaux sont alors comparés en calculant la similarité de leur histogramme de succession de notes.

Publication

E. DELLANDREA, H. HARB, L. CHEN. Zipf, neural networks and svm for musical genre classification. IEEE International Symposium on Signal Processing and Information Technology (ISSPIT 2005). Athènes (Grèce). p. 57-62. Décembre 2005.

A.Paradzinets, H.Harb, L.Chen, Use of Continuous Wavelet-Like Transform in Automated Music Transcription, *14th European Signal Processing Conference (EUSIPCO06)* (2006), Florence, Italy, *accepted*