

MUSICDISCOVER

Introduction

1. Tache 5. Recherche de la musique par similarité

Dans ce projet nous avons étudié la similarité musicale en analysant le signal audio. La similarité musicale est la mesure qui permet à un sujet humain de qualifier de similaire ou non-similaire deux segments musicaux. Idéalement, une mesure de similarité musicale devrait être capable de qualifier comme très similaire l'ensemble des interprétations par différents auteurs de différents genres musicaux d'une même chanson.

La plupart des techniques existantes sont basées sur la modélisation des coefficients cepstrales. Les caractéristiques les plus connues de la famille des caractéristiques cepstrales sont les Mel Frequency Cepstral Coefficients (MFCC) qui sont des coefficients de cepstre obtenus des spectres de fréquence filtrés par une banque de filtres conformément à l'échelle de MEL. L'échelle de MEL est une échelle reflétant la perception humaine avec des différentes précisions fréquentielles suivant les fréquences. Pour simplifier, l'échelle de MEL affirme que les hautes fréquences sont aperçues avec moins de précision que les basses fréquences. Cependant, si ce modèle est convenable pour l'analyse de la parole et de ce fait il est à la base de systèmes de reconnaissance automatique de la parole, il n'est pas nécessairement approprié pour l'analyse d'un signal sonore en général, et d'un signal musical en particulier.

La ligne conductrice de nos travaux est de compléter les techniques classiques basées sur MFCC avec des mesures de similarité sur des caractéristiques musicales.

1.1.Extraction des caractéristiques

Nos principales contributions aux différentes sous-tâches dans le cadre du projet MusicDiscover sont les suivantes :

- ✓ Elaboration des algorithmes d'analyse sonore adaptés à la musique.

Le signal musical est un signal portant des caractéristiques

spécifiques. Il est quasi-stationnaire, les fréquences ont une échelle logarithmique qui rend l'utilisation des techniques standard comme MFCC moins appropriées à l'analyse du signal musical. Une transformation inspirée par la transformation en ondelette permettant d'obtenir les échelles tempo-spectrales adaptées à la musique a été proposée. Cette transformation à l'échelle variable (Variable Resolution Transform – VRT) est une transformation qui hérite des caractéristiques de transformation en ondelettes classique en améliorant la résolution fréquentielle dans la zone des hautes fréquences. Cette modification permet de mieux récupérer les harmoniques du signal.

- ✓ Développement d'un algorithme de détection de coups de batterie et des mesures de similarité.

Ici, un algorithme de « beat detection » a été proposé. L'algorithme est basé sur la transformation VRT et les techniques de traitement d'images. Cet algorithme produit une courbe de probabilité de « beat ». Le traitement suivant construit à partir de cette courbe une nouvelle forme de représentation rythmique qui est l'histogramme de beat en 2D. L'algorithme permet également d'estimer le tempo précis d'une pièce musicale à partir de l'histogramme 2D. La comparaison de performance d'estimation du tempo avec les techniques existantes a montrée les résultats proches, voir meilleurs selon le cas.

- ✓ Elaboration d'algorithme de détection de notes et des mesures de similarité.

Dans cette partie de travail, une approche de détection de notes à partir du signal audio a été propose [1]. Cette approche utilise la transformation VRT comme celle qui est utilisée dans toutes les autres parties du traitement. L'algorithme de détection fait la modélisation du spectre afin de trouver les structures harmoniques correspondantes aux notes. L'information sur les notes permet de construire les mesures de similarité mélodiques telles qu'un histogramme de succession de notes, tonalité et timbre [2].

1.2.Vérifaction

1.2.1.Recherche des titres musicaux avec plusieurs interprétations.

Ici, nous avons proposé d'injecter une liste de titres musicaux avec plusieurs interprétations dans une base de données de musique. Les essais suivants devaient produire des playlists similaires pour chacun des titres. Ensuite, les positions des autres interprétations du même titre ont été surveillées. Cette expérimentation a démontré les résultats corrects

1.2.2.Evaluation de composition de playlists avec avis d'utilisateurs (user feedback).

Cette partie d'expérimentation consistait à collecter les votes d'utilisateurs afin d'évaluer la performance des mesures de similarité. L'idée de données « ground truth » était de proposer les playlists similaires avec insertion de titres de manière aléatoire. La suite de la procédure consistait à comparer les histogrammes de votes d'utilisateurs pour les « vraies » similarités et les similarités fausses. Cette expérimentation a montré une différence statistiquement significative entre les distributions de votes pour les vraies et les fausses similarités. Dans le cadre de ce travail, plusieurs méthodes de combinaison de distances ont été proposées.

2. Tache 6. Reconnaissance de titres musicaux

L'objectif de l'identification de musique dans le cadre du projet est d'intégrer un identifiant unique à chaque titre. Un identifiant obtenu à partir d'une signature du signal sonore aura l'avantage d'être unique et de ne pas tenir compte des erreurs possibles des métadonnées (nom du titre, artiste) ou des multiples versions du même titre. Nous avons testé l'utilisation des probabilités d'appartenance à des classes de musique prédéfinies obtenues par un classificateur généraliste. Les premières expérimentations ont montré que ce type d'identifiant n'est pas suffisamment discriminant pour une base de donnée de plusieurs milliers de titres musicaux. Cet identifiant a cependant l'avantage d'être très compacte. En conséquence nous avons adapté l'ADN sonore que nous avons proposée ultérieurement à l'identification des titres dans le cadre du projet. L'ADN sonore, une chaîne de caractères, consiste en une modélisation de la structure temporelle et fréquentielle du signal sonore.

L'algorithme développé pour l'extraction de l'ADN sonore consiste en

- 1- une segmentation temporelle et fréquentielle du spectre du signal
- 2- estimation de similarité entre les fenêtres consécutives
- 3- codage des valeurs de similarité par des caractères
- 4- constitution de la signature à partir des caractères

Dans le cadre du projet la technique basée sur l'ADN sonore a été réalisée en système automatique de reconnaissance. Le système consiste d'un serveur téléphonique permettant d'enregistrer les morceaux musicaux pour faire la reconnaissance, et le d'un serveur de reconnaissance.

3. Tache 7. Classification de titres musicaux en genre et en émotion

3.1.1. Classification en genre

Dans le cadre de cette tache nous avons créé une base de données de référence pour la classification en genres. Cette base comprend environ 1800 titres musicaux classés en 6 genres. Cette base de référence mieux équilibrée par rapport aux bases existantes nous a permis de réaliser les tests de base pour vérifier la validité des mesures de similarité. La méthode que nous avons utilisée est un classificateur kNN appliqué sur les distances de similarité musicales. Nous avons démontré également, que une application des caractéristiques musicales en combinaison avec les caractéristiques basique spectrales permet d'améliorer les taux de classification à une manière significative.

3.1.2. Classification en émotions

Les descripteurs de la taxonomie d'émotions musicales sont basés sur la théorie général des émotions et comprend deux moyens de description : discrète et dimensionnelle. Hevner et Farnsworth ont proposé une liste des adjectives décrivant les émotions musicales dan les groups différents [3][4]. En 1990s, Thayer [5] a proposé une modèle 2-dimensionnelle qui décrit les émotions musicales avec deux facteurs : la dimension de stress (happy/anxoius) et la dimension d'énergie (calm/energetic), et divise les émotions musicales en 4 quadrants dans l'espace 2-D : "Contentment", "Depression", "Exuberance" et

“Anxious/Frantic”.

Dans nos travaux nous avons proposé deux approches de classification automatique de la musique en émotions basées sur la fusion des caractéristiques avec la théorie de l'évidence : l'approche hiérarchique avec jugement définie et une approche ambiguë avec jugements multiples. En appliquant la modèle de Thayer et une base de données de 607 samples de 8 seconds dérivée de la musique classique, un taux de classification correcte de $83.82 \pm 2.15\%$ a été atteint.

Références

- [1] Paradzinets A., Harb H., Chen L., “Use of Continuous Wavelet-like Transform in Automated Music Transcription”, /Proceedings of EUSIPCO /2006

- [2] A Paradzinets, O. Kotov, H Harb, L. Chen., Continuous Wavelet-like Transform Based Music Similarity Features for Intelligent Music Navigation. /Proceedings of CBMI07, Bordeaux, France. / 2007

- [3] Hevner, K.: Experimental studies of the elements of expression in music. American Journal of Psychology 48 (1936) p246–268

- [4] Paul R. Farnsworth. The social psychology of music. The Dryden Press, 1958

- [5] Thayer, R. E., The biopsychology of mood and arousal, Oxford University Press, 1989