

شناسایی خودکار سبک موسیقی

حامد حبیبی اقدم

محمد مهدی همایون پور

دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر

{habibiaghdam, homayoun}@aut.ac.ir

چکیده

در دهه اخیر به دلیل رشد سریع داده‌های موسیقی که به صورت فایل‌های صوتی در اینترنت و از طریق پایگاه داده‌های بسیار بزرگ قابل دسترس هستند، توجه محققین به روش‌های پردازش خودکار سیگنال‌های موسیقی بیشتر از گذشته شده است. شناسایی خودکار سبک موسیقی نیز به عنوان یکی از زمینه‌های تحقیقی جذاب در این عرصه می‌باشد. در این مقاله روش‌های مختلفی برای استخراج ویژگی، انتخاب ویژگی و مدل‌سازی برای شناسایی خودکار ۸ سبک موسیقی شامل سبک‌های *Metal Jazz*, *Classic Piano*, *Classic*, *Celtic*, *Persian Classic*, *Relaxing* و *Dance* که از آلبوم‌هایی با نوازنده‌های مختلف جمع‌آوری شدند، پیاده‌سازی و مورد ارزیابی قرار گرفت. از میان روش‌های رایج استخراج ویژگی که در حوزه تشخیص گفتار، شناسایی آلت موسیقی و همچنین شناسایی سبک موسیقی مورد استفاده قرار می‌گیرند، انواع ویژگی‌های کوتاه مدت، میان مدت و بلند مدت بررسی شده و از ویژگی‌های بلند مدت به دلیل اینکه در تعیین سبک موسیقی کارایی چندانی نداشتند، صرف نظر گردید. با استفاده از یک روش انتخاب ویژگی و مطالعه ترکیب‌های مختلف ویژگی‌ها، بردار ویژگی مناسب به دست آمد که نهایتاً با استفاده از حدود ۳۰ ویژگی برتر انتخاب شده از بین ۱۲۲ ویژگی، به کارایی ۹۰٪ در تشخیص سبک موسیقی دست یافتیم. مدل‌سازی سبک‌های موسیقی با و بدون در نظر گرفتن توالی زمانی مورد ارزیابی قرار گرفت. از روش‌های مدل‌سازی مانند شبکه‌های عصبی، مدل مخلوط گاوسی، درخت تصمیم، ماشین بردار پشتیبان به تنهایی و نیز در یک ساختار سلسله‌مراتبی که در آن شناسایی سبک موسیقی از کلاس‌های با تعداد بیشتری سبک موسیقی شروع و به کلاس‌های با تنها یک سبک موسیقی ختم می‌گردد استفاده شد. کارایی بدست آمده با استفاده از این ساختار سلسله‌مراتبی ۸۸٪ و در حدود ۲٪ از ساختار مسطح کمتر است، ضمن اینکه ساختار سلسله‌مراتبی برای تعیین سبک موسیقی در مواردی که با تعداد زیادی از سبک‌های موسیقی مواجه هستیم دلیل نیاز به استفاده از تعداد کمتری طبقه‌بندی کننده در مرحله آزمایش، از سرعت بالاتری برخوردار است. علاوه بر این در یک ساختار سلسله‌مراتبی دیگر، سعی شد که دسته‌بندی کننده‌ها به صورت ترکیب موازی مورد استفاده قرار گرفته و تصمیم‌گیری بر اساس رای‌گیری نسبت به نتیجه به دست آمده از دسته‌بندی کننده‌ها انجام شود.

کلمات کلیدی: شناسایی سبک موسیقی، ویژگی طیفی، کنتراست طیفی مبتنی بر معیار اکتاو، شدت سیگنال مبتنی بر معیار اکتاو، ضرائب کسپترال مبتنی بر معیار مل، تجمیع ویژگی

۱ - مفاهیم اولیه

موسیقی تنها به عنوان هنری برای ابراز احساسات نمی‌باشد و امروزه به دلیل تاثیر اجتماعی آن، از موسیقی برای اهداف مختلف استفاده می‌شود. با آغاز قرن ۲۱م، حجم اطلاعات موسیقی در سطح اینترنت با رشد سریعی مواجه شد و اکنون با پایگاه اطلاعات عظیمی از موسیقی‌هایی با سبک‌های مختلف در اینترنت مواجه هستیم. همین امر باعث شده است تا نیاز به روش‌های پردازش دیجیتالی سیگنال‌های موسیقی بیشتر از گذشته احساس شود. امروزه تحقیقات گسترده‌ای در زمینه‌های بازیابی موسیقی از پایگاه داده‌های بزرگ، شناسایی آلت‌های موسیقی، استخراج ریتم و آهنگ موسیقی و موارد مشابه دیگر انجام می‌پذیرد.

در این بین شناسایی سبک موسیقی از محبوبیت بیشتری برخوردار بوده و به دلیل پیچیدگی‌ها و دسته‌های بسیاری که در این مساله وجود دارد، تا کنون روش کارآمدی برای شناسایی مجموعه بزرگی از سبک‌های موسیقی در عمل ارائه نشده است. توسعه سیستمی که بتواند به طور خودکار سبک فایل موسیقی را شناسایی کند، نیاز به انجام دو مرحله دارد. در مرحله اول سعی می‌کنیم فایل صوتی را به شکل مجموعه‌ای از بردارهای ویژگی نشان دهیم. سپس در مرحله دوم از همین بردارهای ویژگی استفاده کرده و مدلسازی هریک از سبک‌های موسیقی را انجام می‌دهیم.

تا کنون مطالعات زیادی برای استخراج ویژگی از فایل صوتی صورت گرفته است که نتیجه آن معرفی روش‌های مختلفی برای به دست آوردن بردارهای ویژگی فایل‌های صوتی می‌باشد. بردارهای ویژگی فایل صوتی را از نظر زمانی می‌توان به سه دسته تقسیم کرد: (۱) ویژگی‌های کوتاه‌مدت که معمولا در بازه‌های ۲۰ الی ۴۰ میلی‌ثانیه استخراج

می‌شوند. چراکه در این بازه زمانی می‌توان ویژگی‌های آماری سیگنال ثابت فرض کرد. (۲) ویژگی‌های میان‌مدت که در بازه‌های ۱ الی ۳ ثانیه‌ای و با تجمیع^۱ بردارهای ویژگی کوتاه‌مدت همسایه به دست می‌آیند. (۳) ویژگی‌های بلند مدت که معمولا با استفاده از نمونه‌های موجود در کل سیگنال موسیقی محاسبه می‌شوند. ویژگی‌های کوتاه‌مدت و میان‌مدت را به طور کامل در این مقاله بررسی می‌کنیم. ویژگی‌های بلند مدت همانند ویژگی‌های مبتنی بر ریتم [۱] و هیستوگرام گام^۲ [۱۰] همانگونه که در [۱] نیز نشان داده شده است، تاثیر کمی در افزایش دقت شناسایی سبک موسیقی دارند. به همین دلیل از بررسی این نوع ویژگی‌ها صرف نظر می‌کنیم. روش‌های مدلسازی سبک موسیقی را نیز می‌توان به دو دسته تقسیم کرد:

- روش‌هایی که به طور صریح ترتیب وقوع بردارهای ویژگی را در نظر نمی‌گیرند.
- روش‌هایی که ترتیب وقوع بردارهای ویژگی در آن‌ها دارای اهمیت می‌باشد.

روش‌های دسته‌بندی اول شامل شبکه‌های عصبی، ماشین بردار پشتیبان، نزدیکترین همسایه، درخت تصمیم و مدل مخلوط گاوسی بوده و روش‌های دسته‌بندی دوم نیز شامل مدل مخفی مارکوف و روش‌های مبتنی بر احتمال بیزی همانند ماتریس هم‌پیشامد^۳ یا چندتایی‌ها^۴ می‌باشد.

آقای Tzanetakis و همکارشان [۱] با استفاده ضرائب کپسترال مبتنی بر معیار مل، ویژگی‌های طیفی و همچنین اطلاعات ریتم و گام موسیقی به شناسایی ۱۰ سبک مختلف پرداختند. ایشان در این مقاله علاوه بر معرفی روش استخراج

¹ Integrate

² Pitch

³ Co-occurrence

⁴ N-Gram

ریتم موسیقی، نشان داده‌اند که این اطلاعات تاثیر کمی در افزایش قدرت دسته‌بندی دارند. [۲] نیز بدین منظور ابتدا یک فایل موسیقی را به سه بخش ۳۰ ثانیه‌ای تقسیم کرده و سپس برای هر بخش بردار ویژگی مشابهی را استخراج نموده اند (به این کار تفکیف زمانی^۱ نیز می‌گویند). سپس هریک از این سه قسمت را به صورت جداگانه دسته‌بندی کرده و در نهایت با استفاده از روش آرا و همچنین روش احتمالی و از روی اطلاعات به دست آمده از این سه بخش، ۹ سبک موسیقی مختلف را شناسایی کرده‌اند. ایشان از روش‌های دسته‌بندی درخت تصمیم، ۳ تا نزدیکترین همسایه، شبکه عصبی، بیز ساده و ماشین بردار پشتیبان برای مدل‌سازی سبک‌های موسیقی استفاده کرده‌اند.

نویسندگان [۳] در مقاله خود، دو ویژگی بلند مدت جدید به نام‌های **MSFM**^۲ و **MSCM**^۳ پیشنهاد داده‌اند. [۴] نیز از بعد دیگری به مدل‌سازی این مساله پرداخته و سعی کرده است با در نظر گرفتن توالی رخداد بردارهای ویژگی، وابستگی بین آن‌ها را نیز مدل کند. روشی که در این مقاله ارائه شده است در نهایت هریک از سبک‌های موسیقی را به صورت یک دوتایی^۴ مدل‌سازی می‌کند. سپس با مشاهده ترتیبی از بردارهای ویژگی و با محاسبه لگاریتم احتمال مشاهدات، میزان محتمل بودن بردار مشاهدات را به ازای هریک از مدل‌ها به دست آورده و محتمل‌ترین مدل را به عنوان پاسخ نهایی بر می‌گرداند. علاوه بر آن، این مقاله روش ارائه شده را با روش مدل مخفی مارکوف چپ به راست و ارگودیک مقایسه کرده است. با این حال روش ارائه شده در این مقاله به دلیل اینکه از دوتایی‌های گسسته استفاده می‌کند، دارای نتایج ضعیفتری نسبت به سایر روش‌ها می‌باشد. همانطور که نویسندگان مقاله نیز به آن اشاره کرده‌اند، به احتمال زیاد می‌توان با مدل‌سازی پیوسته دوتایی‌ها دقت سیستم را افزایش داد.

نویسندگان [۵] از شبکه‌های عصبی تاخیر زمانی^۵ برای شناسایی ۸ سبک مختلف استفاده کرده‌اند و [۱۲] نیز پس از معرفی روش استخراج ویژگی ضرایب هیستوگرام موجک داب^۶ به مقایسه روش‌های استخراج ویژگی مختلف پرداخته و با اعمال آن برای دسته‌بندی ۱۰ سبک موسیقی، نشان داده است که نتایج به دست آمده با استفاده از این روش استخراج ویژگی قابل مقایسه با سایر روش‌های مرسوم می‌باشد. [۶] نیز روش استخراج ویژگی‌های میان‌مدت **کنتراست طیف مودولاسیون مبتنی بر معیار اکتاو**^۷ را **پیشنهاد** داده و آن را با روش کنتراست طیفی مبتنی بر معیار اکتاو^۸ و ضرایب کسپترال مبتنی بر معیار مل مقایسه کرده است. نتایج گزارش شده در مقاله‌های مختلف نشان داده است که سه روش آخر که یادآور شدیم، از توانایی قابل قبولی برای حل مساله شناسایی سبک موسیقی برخوردار هستند. در نهایت [۱۵] روش‌های مختلف تجمیع بردارهای ویژگی و به دست آوردن بردارهای ویژگی میان‌مدت را مورد بررسی قرار داده و آن‌ها را با استفاده از شناسایی ۱۱ سبک موسیقی مقایسه کرده است.

در ادامه، ما نیز ابتدا روش‌های مختلف مدل‌سازی مسئله را در بخش ۲ نشان داده و سپس در بخش‌های ۳ و ۴ به ترتیب روش دسته‌بندی سبک‌های موسیقی و روش‌های استخراج ویژگی را مورد بررسی قرار می‌دهیم. همچنین در بخش ۵ نشان می‌دهیم که چگونه می‌توان با استفاده از روش‌های تجمیع، ویژگی‌های میان‌مدت سیگنال موسیقی را به دست آورد. فصل ۶ نیز نتایج تجربی به دست آمده را نشان و بخش ۷ یک جمع‌بندی از مباحث را ارائه می‌کند. در نهایت بخش ۸ به نتیجه‌گیری از این تحقیق می‌پردازد.

۲- روش‌های مدل‌سازی

⁵ TDNN : Time-Delay Neural Network

⁶ DWCH : Daubechies Wavelet Coefficients

Histogram

⁷ OMSC : Octave-Based Modulation Spectral Contrast

Contrast

⁸ OSC : Octave-Based Spectral Contrast

¹ Time Decomposition

² Modulation Spectral Flatness Measure

³ Modulation Spectral Crest Measure

⁴ Bigram

محققان در مطالعات خود روند شناسایی سبک موسیقی را به شیوه‌های مختلفی انجام داده‌اند [۳]. روش اول که در شکل ۱ نشان داده شده است، بدین صورت است که ابتدا سیگنال موسیقی را به قاب‌های ۲۰-۴۰ میلی‌ثانیه همپوشان تقسیم کرده و بردار ویژگی هریک از قاب‌ها را جداگانه به دست می‌آوریم. به قاب‌های ۲۰-۴۰ میلی‌ثانیه‌ای در اصطلاح پنجره تحلیلی^۱ می‌گویند. بعد از آنکه بردار ویژگی همه پنجره‌های تحلیلی (ویژگی‌های کوتاه‌مدت) به دست آمدند، سعی می‌کنیم با استفاده از بردارهای ویژگی استخراج شده از چند پنجره تحلیلی پشت سرهم (مثلاً ۴۰ پنجره تحلیلی)، یک بردار ویژگی به دست آوریم که بتواند اطلاعات آماری پنجره‌های تحلیلی را به شکل مجتمع نشان دهد. به تعداد پنجره‌های تحلیلی که برای به دست آوردن ویژگی میان‌مدت سیگنال موسیقی استفاده می‌کنیم، در اصطلاح پنجره بافت^۲ بافت^۲ می‌گویند. ویژگی‌های میان‌مدت با تجمیع ویژگی‌های کوتاه‌مدت انجام می‌پذیرند که روش‌های آن را در ادامه مورد بررسی قرار داده‌ایم. سپس سعی می‌کنیم با تجمیع همه بردارهای ویژگی میان‌مدت به دست آمده، یک بردار ویژگی به دست آورده و آن را پس از اعمال به الگوریتم دسته‌بندی کننده، خروجی نهایی را به دست آوریم.

روش دوم (شکل ۲) برای شناسایی سبک موسیقی نیز بدین صورت است که همانند روش اول، ابتدا ویژگی‌های کوتاه‌مدت و میان‌مدت سیگنال را به دست می‌آوریم. سپس هریک از ویژگی‌های میان‌مدت را دسته‌بندی کرده و خروجی نهایی را با اعمال یک الگوریتم انتخاب مبتنی بر رای‌گیری^۳ بر روی خروجی به دست آمده از دسته‌بندی تک تک ویژگی‌های میان‌مدت، به دست آوریم.

در دو روش اول توالی بردارهای ویژگی به طور صریح مورد توجه قرار نمی‌گیرد، بلکه روش تجمیع بردارهای ویژگی کوتاه‌مدت و بردار ویژگی میان‌مدت به دست آمده از آن است که به طور ضمنی این توالی را مدلسازی می‌نماید.

روش سوم (شکل ۳) شناسایی سبک موسیقی با دو روش اول متفاوت بوده و در آن توالی مشاهده بردارهای ویژگی به طور مستقیم مورد توجه قرار می‌گیرد.

در این روش [۴] - در مرحله آموزش- ابتدا ویژگی‌های کوتاه‌مدت هریک از فایل‌های موسیقی به دست می‌آیند و سپس یک الگوریتم خوشه‌بندی بر روی بردارهای ویژگی کوتاه‌مدت هریک از فایل‌های موسیقی اعمال می‌شود. در این مرحله به ازای هر فایل موسیقی، بردارهای ویژگی کوتاه‌مدت آن را به k_1 خوشه تقسیم می‌کنیم. با فرض اینکه n فایل موسیقی نیز داریم، می‌توان نتیجه گرفت که در پایان مرحله اول، در مجموع $k_1 \times n$ مرکز خوشه خواهیم داشت. یکبار دیگر نیز الگوریتم خوشه‌بندی را بر روی این تعداد مرکز خوشه اعمال کرده و آن‌ها را به k_2 خوشه تقسیم می‌کنیم. به k_2 خوشه به دست آمده می‌توان به عنوان یک لغتنامه با k_2 مدخل نگریست. بعد از آن که این لغتنامه از همه فایل‌های موسیقی به دست آمد، دوتایی‌های (یا چندتایی^۴) هریک از سبک‌های موسیقی را تشکیل می‌دهیم. در مرحله آزمایش نیز سعی می‌کنیم احتمال رخداد توالی بردارهای ویژگی را با استفاده از روش بیزی پیدا کنیم. جزئیات بیشتر این روش را می‌توانید در [۴] مشاهده کنید. علت خوشه‌بندی دو سطحی بردارهای ویژگی کوتاه‌مدت در این روش را می‌توان به دست آوردن احتمال شرطی گسسته آن‌ها دانست. با این حال به نظر می‌رسد استفاده از احتمال شرطی پیوسته توالی بردارهای ویژگی ممکن است موجب افزایش دقت این روش گردد. روش سوم علاوه بر اینکه به طور مستقیم توالی بردارهای ویژگی را در نظر می‌گیرد، مزیت دیگری نیز دارد که به لحاظ تئوری از اهمیت ویژه‌ای برخوردار است.

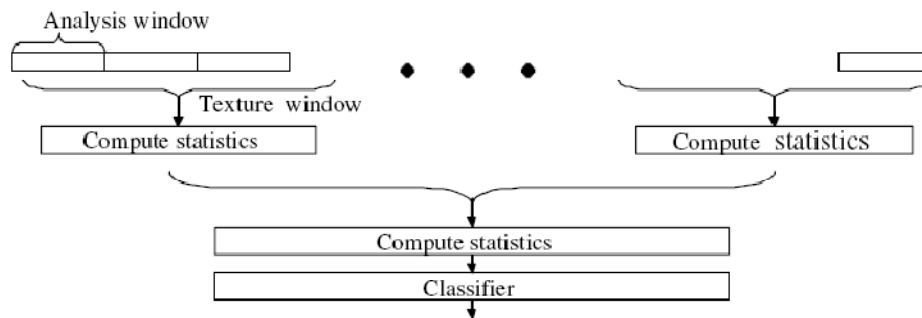
احتمال اینکه همه قسمت‌های یک فایل موسیقی قطعاً متعلق به یک سبک خاص باشد، بسیار پایین است. این قضیه در مورد سبک‌های موسیقی همچون سبک Celtic، Folk، Pop و موارد مشابه دیگر پر رنگ‌تر نیز می‌شود.

¹ Analysis Windows

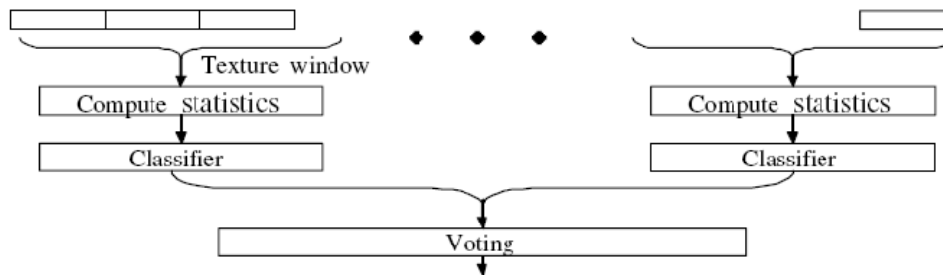
² Texture Window

³ Voting Algorithm

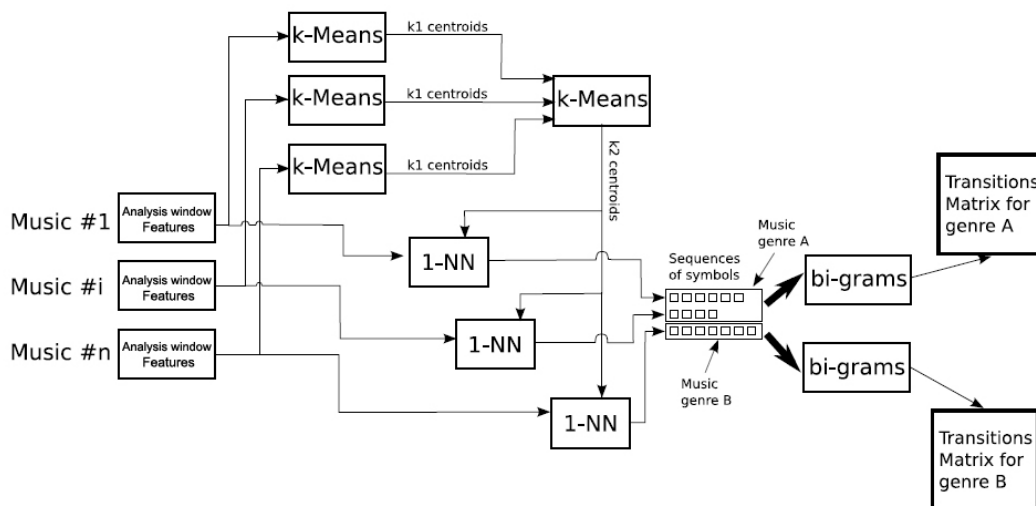
⁴ N-Gram



شکل ۱: شناسایی سبک موسیقی با تجمیع بردارهای ویژگی کوتاه‌مدت و میان‌مدت^۱



شکل ۲: شناسایی سبک موسیقی با دسته‌بندی ویژگی‌های میان‌مدت و الگوریتم انتخاب آرا^۲



شکل ۳: شناسایی سبک موسیقی با دوتایی‌ها [۴]

^۱ (Dalwon Jang 2008)

در مرحله آموزش دو روش اول، نیاز به تعداد زیادی نمونه آموزشی داریم که دسته‌های آن‌ها مشخص باشند. به عنوان مثال سبک Celtic را در نظر بگیرید. در دو روش اول مدلسازی، برای جمع‌آوری داده‌های آموزشی، تعدادی فایل موسیقی که بخش زیادی از آن متعلق به سبک Celtic می‌باشند را جمع‌آوری کرده و هنگام برچسب‌گذاری پنجره‌های بافت، برچسب همه آن‌ها را Celtic در نظر می‌گیرند. در اکثر موارد همه بخش‌های یک موسیقی Celtic، دارای سبک یکسانی نیستند و ممکن است چند ثانیه از همین موسیقی به سبک Pop یا Classic نواخته شود. این بدان معناست که با این کار حتی بخش‌هایی از موسیقی که سبک آنها Celtic نمی‌باشد، با این برچسب نشانه‌گذاری کرده‌ایم و به عبارت دیگر موجب تولید داده‌های غلط در مجموعه داده آموزشی شده‌ایم. به راحتی می‌توان نشان داد که در این روش داده‌های متناقض نیز تولید می‌شوند. تنها روشی که می‌تواند به طور کامل این مسئله را حل کند، اینست که مدت زمان زیادی را صرف کرده و تنها بخش‌هایی از هر فایل موسیقی را انتخاب کنیم که سبک آن‌ها دقیقاً Celtic می‌باشد. بدیهی است چنین کاری هنگام پیاده‌سازی در عمل امکان‌پذیر نمی‌باشد.

ولی در روش سوم از ابتدا برچسب قسمت‌های مختلف را تعیین نمی‌کنیم، بلکه نخست یک لغتنامه از نمادها تشکیل داده و سپس هریک از سبک‌های فایل‌های موسیقی را جداگانه بررسی کرده و در آن، انتقال بین بردارهای ویژگی را مدنظر قرار می‌دهیم. در هر دو سبک Celtic و Rock می‌توان بخش‌هایی را پیدا کرد که به سبک Pop نواخته شده‌اند. این در حالی است که دو روش قبلی موجب تولید داده‌های آموزشی متناقض در این مورد می‌شوند. ولی روش سوم انتقال از یک سبک به سبک دیگر را مدل می‌کند. از اینرو احتمال وجود داده متناقض یا داده اشتباه در آن بسیار پایین‌تر خواهد بود. توجه به این نکته حائز اهمیت است که این مشکل بیشتر مبنای نظری داشته و در عمل زمانی که از داده‌های آموزش زیادی استفاده می‌کنیم، با مشکل چندانی در این رابطه مواجه نخواهیم شد.

روش چهارم که شکل دیگری از روش سوم است، بدین شکل عمل می‌کند که بجای خوشه‌بندی دو سطحی فایل‌های موسیقی و ایجاد لغتنامه از بردارهای ویژگی، هریک از سبک‌های موسیقی را به صورت توابع مخلوط گاوسی شرطی مدلسازی نموده و هنگام آزمایش، محتمل‌ترین مدل را به عنوان نتیجه نهایی برگرداند.

روش پنجم استفاده از مدل مخفی مارکوف ارگودیک می‌باشد [۴] که در آن سعی می‌کنیم به ازای هریک از سبک‌های موسیقی، یک مدل مخفی مارکوف به دست آوریم. روش ششم استفاده از شبکه‌های عصبی تاخیر زمانی است که این شبکه‌ها به خوبی قادر به مدلسازی توالی زمانی بردارهای ویژگی هستند [۵]. فایل‌های موسیقی بسیاری وجود دارند که ترکیبی از چند سبک موسیقی می‌باشند. به عنوان مثال فایل موسیقی Avemano Orchestral از Era ابتدا با سبک موسیقی Gospel شروع شده و سپس به سبک Pop تبدیل می‌شود. سپس در اواسط موسیقی، سبک آن به Rock تغییر پیدا کرده و مجدداً به سبک Pop باز می‌گردد. بنابراین ممکن است نیاز داشته باشیم که میزان وجود هر یک از سبک‌های موسیقی در یک آهنگ را مشخص نماییم. در این خصوص می‌توانیم از روش‌های مدلسازی دوم تا پنجم استفاده کنیم.

۳- نحوه دسته‌بندی سبک‌های موسیقی

نکته‌ای که در خصوص مساله شناسایی سبک موسیقی وجود دارد اینست که در این مساله با تعدادی بسیار زیادی از سبک‌های موسیقی مواجه هستیم. بیشتر روش‌هایی که تا کنون ارائه شده است، از یک روش دسته‌بندی مسطح^۱ استفاده کرده‌اند [۸]. به عنوان مثال با فرض ۲۰ سبک موسیقی، روش دسته‌بندی مسطح بدین شکل عمل می‌کند که همه ۲۰ سبک مذکور را هنگام انجام عمل دسته‌بندی بررسی می‌کند. بدیهی است که این روش با افزایش تعداد

¹ Flat Classification

دسته‌ها کارآمد نخواهد بود. روش دیگری که در سال‌های اخیر مورد توجه محققین قرار گرفته است، استفاده از روش دسته‌بندی سلسله مراتبی است. با فرض ۲۱ سبک موسیقی، روش سلسله مراتبی ابتدا یک درخت سلسله مراتبی شامل ۲۱ برگ تشکیل می‌دهد. فرض کنید درخت تشکیل شده دارای دو سطح بوده که در سطح اول آن ۳ گره وجود دارد. هریک از این گره‌ها نیز دارای ۷ برگ هستند. بدین ترتیب درختی دو سطحی با ۲۱ گره در دست داریم. هر برگ در این درخت نشان دهنده یک سبک موسیقی است. حال زمانی که می‌خواهیم سبک یک فایل موسیقی را شناسایی کنیم، به جای مقایسه با ۲۱ دسته مختلف، ابتدا سیگنال صوتی را با ۳ گره اول درخت مقایسه کرده و سپس برگ‌های گره برنده در سطح اول را مورد بررسی قرار می‌دهیم.

تعیین سلسله مراتب سبک‌های موسیقی در بیشتر موارد امری دلخواه می‌باشد. با این حال سلسله مراتب‌هایی که در سایت‌هایی همانند Amazon، Allmusic و سایر سایت‌های مشابه انجام شده‌اند می‌تواند به عنوان نقطه شروع برای حل مساله شناسایی خودکار سبک موسیقی به کار رود. آقای تائولی و همکارانش نیز در [۸] با استفاده از ماتریس تداخل، روشی را برای به دست آوردن سلسله مراتب سبک‌های موسیقی معرفی کرده‌اند. استفاده از روش سلسله مراتبی به این دلایل مفید خواهد بود که: (۱) بیشتر کاربران به جای استفاده از پرس‌وجوهای کلی، ترجیح می‌دهند اطلاعات موسیقی را به صورت سلسله مراتبی مورد جستجو قرار دهند (۲) استفاده از ساختار سلسله مراتبی موجب شناسایی وابستگی‌های بین سبک‌های موسیقی شده و اطلاعات مفیدی را در اختیار ما قرار می‌دهد. (۳) استفاده از ساختار سلسله مراتبی موجب کاهش خطای سیستم و قابل پذیرش شدن آن می‌شود.

۴- روش‌های استخراج ویژگی

در این بخش به بررسی روش‌های مختلف استخراج ویژگی معرفی شده برای شناسایی سبک موسیقی را مورد بررسی قرار داده و برخی از ویژگی‌هایی که در سایر مسائل مربوط به

موسیقی همانند شناسایی آلت موسیقی کاربرد داشته و برای حل مساله شناسایی سبک موسیقی نیز می‌توانند بکار روند، معرفی می‌کنیم. ویژگی‌های دیگری همانند توصیف گره‌های MPEG-7 [۹]، ریتم موسیقی [۱] و هیستوگرام گام چندتایی [۱۰] نیز وجود دارند که در این تحقیق ما از آن‌ها استفاده نکرده‌ایم. برای استخراج ویژگی‌هایی که در ادامه می‌آیند، سیگنال موسیقی رقمی شده (با نرخ نمونه‌برداری ۴۴۱۰۰ نمونه در ثانیه با دقت ۱۶ بیت به ازای هر نمونه) را به پنجره‌های ۴۰ میلی‌ثانیه (فرض می‌کنیم در این بازه خصوصیات اماری سیگنال ثابت است) با همپوشانی ۲۰ میلی‌ثانیه تقسیم کرده و سپس به ازای هر قاب ویژگی‌های زیر را استخراج کرده‌ایم.

۴-۱- ویژگی‌های زمانی^۱ سیگنال

ویژگی‌های زمانی از روی نمونه‌های سیگنال در حوزه زمان به دست می‌آیند که رایج‌ترین آن‌ها انرژی، نرخ عبور از صفر و میزان متناوب بودن سیگنال می‌باشد.

انرژی: انرژی یک قاب برابر است با مجموع توان دوم هریک از نمونه‌ها که به شکل نشان داده شده در رابطه (۱) تعریف می‌شود [۱۹]. که در این رابطه $x(n)$ نشان دهنده مقدار نمونه n در قاب است.

$$\text{energy} = \sum_{n=0}^{N-1} (x[n]^2) \quad (1)$$

نرخ عبور از صفر: این ویژگی نشان‌دهنده تعداد دفعاتی است که سیگنال از مقدار صفر عبور کرده است و به عنوان معیاری برای میزان نویزی بودن سیگنال به حساب می‌آید. این ویژگی را می‌توان به صورت زیر تعریف کرد [۱۹]. در این رابطه تابع sign برای مقادیر مثبت مقدار ۱ و برای مقادیر منفی مقدار ۰ برمی‌گرداند.

$$\text{zcr} = \frac{1}{N} \sum_{n=0}^{N-1} |\text{sign}(x[n]) - \text{sign}(x[n-1])| \quad (2)$$

¹ Temporal

$$bass = \sqrt{\frac{\sum_{k=0}^{N/2} ramp(k) \times mag(X[k])^2}{\sum_{k=0}^{framesize/2} ramp(k)}} \quad (5)$$

مرکز ثقل طیفی (میزان شفافیت): برای محاسبه میزان شفافیت سیگنال می‌توان از رابطه (۶) استفاده کرد [۱۹]. در این رابطه تابع $freq(k)$ نشان دهنده فرکانس متناظر نمونه k ام در طیف فوریه است.

$$centroid = \frac{\sum_{k=0}^{N/2} freq(k) \times mag(X[k])}{\sum_{k=0}^{N/2} mag(X[k])} \quad (6)$$

پهنای باند (بازه طیفی): این ویژگی محدوده فرکانسی قاب را نشان می‌دهد [۱۹].

$$bandwidth = \frac{\sum_{k=0}^{N/2} |freq(k) - centroid| \times mag(X[k])}{\sum_{k=0}^{N/2} mag(X[k])} \quad (7)$$

جریان طیفی: این ویژگی میزان تغییرات محلی طیف سیگنال را نشان می‌دهد که با استفاده از رابطه (۸) محاسبه می‌شود [۱۹]. در این رابطه $norm[k]$ متناظر با اندازه نرمالیزه مولفه k ام از طیف دامنه فوریه می‌باشد. همچنین f نیز نشان دهنده شماره فریم است.

$$flux = \sum_{k=1}^{N-1} (norm_f[k] - norm_{f-1}[k])^2 \quad (8)$$

۳-۴ ویژگی‌های کپسترال

یکی از مهمترین ویژگی‌های کپسترال که به طور گسترده در مباحث پردازش گفتار استفاده می‌شود، ضرائب کپسترال مبتنی بر معیار مل^۳ است. برای به دست آوردن ضرائب کپسترال یک قاب از سیگنال، ابتدا قاب زمانی در یک پنجره

میزان تناوب: میزان تناوب سیگنال را می‌توان با استفاده از روشی بنام خودهمبستگی^۱ به دست آورد. محل قویترین قله‌ها در نمودار سیگنال خود همبستگی میزان تناوب بودن سیگنال را نشان می‌دهد. برای به دست آوردن مقدار خودهمبستگی سیگنال می‌توان از رابطه (۳) استفاده کرد [۱۹]. در این رابطه مقدار رابطه (۱) را به ازای k از ۰ تا $N-1$ محاسبه کرده و محل قویترین قله‌ها را به عنوان نتیجه بر می‌گردانیم.

$$corr(k) = \sum_{n=0}^{N-1} (x[n]x[n-k]) \quad (3)$$

۲-۴ ویژگی‌های طیفی^۲ سیگنال

برای محاسبه ویژگی‌های طیفی سیگنال در یک قاب، ابتدا تبدیل فوریه نمونه‌های موجود در آن قاب را محاسبه می‌کنیم. همه ویژگی‌های طیفی که در ادامه نشان می‌دهیم با استفاده از نصف مقادیر طیف سیگنال محاسبه می‌شوند در روابطی که در ادامه می‌آیند. $X(k)$ نشان دهنده طیف به دست آمده از تبدیل فوریه سیگنال و $x(n)$ نشان دهنده نمونه سیگنال در حوزه زمان است.

دامنه طیفی: این ویژگی، بلندی سیگنال را نشان می‌دهد که برای محاسبه آن می‌توان از رابطه (۴) استفاده کرد [۱۹]. در این رابطه $mag(\cdot)$ نشان دهنده اندازه نمونه k ام از طیف دامنه (Magnitude Spectrum) فوریه نمونه است.

$$amplitude = \sqrt{\sum_{k=0}^{N/2} mag(x[k])^2} \quad (4)$$

میزان بم بودن: این ویژگی میزان بم بودن سیگنال را نشان داده و با استفاده از رابطه (۵) به دست می‌آید [۱۹]. در این رابطه $ramp$ یک تابع نزولی و خطی بر حسب k است که مقدار آن در $k=0$ برابر با ۱ و مقدار آن در $k=N/2$ برابر با ۰ می‌باشد.

¹ Autocorrelation

² Spectral

³ MFCC: Mel Frequency Cepstral Coefficients

مولفه‌های غیر هامونیک طیفی یا نویز می‌باشد. بنابراین اختلاف بین قله و دره نشان دهنده کنتراست طیفی خواهد بود. نحوه استخراج ویژگی در این روش بدین صورت است که پس از محاسبه طیف قاب، با استفاده از مقادیر نشان داده شده در جدول ۱، آن را به زیرباندهای اکتاو مختلفی تقسیم می‌کنیم.

جدول ۱: نحوه توزیع فیلترها

فیلتر	فرکانس شروع	فرکانس پایان
زیرباند ۱	۰	۱۰۰
زیرباند ۲	۱۰۰	۲۰۰
زیرباند ۳	۲۰۰	۴۰۰
زیرباند ۴	۴۰۰	۸۰۰
زیرباند ۵	۸۰۰	۱۶۰۰
زیرباند ۶	۱۶۰۰	۳۲۰۰
زیرباند ۷	۳۲۰۰	۸۰۰۰
زیرباند ۸	۸۰۰۰	۲۲۰۵۰

فرض کنید $(X_{b,1}, X_{b,2}, \dots, X_{b,N_b})$ نشان دهنده مقادیر طیف سیگنال در زیرباند b ام بوده و N_b نیز تعداد کل مقادیر در این زیرباند را نشان دهد. برای محاسبه کنتراست طیفی مبتنی بر معیار اکتاو، ابتدا این مقادیر را به صورت نزولی مرتب می‌کنیم. فرض کنید مقادیر مرتب شده را به صورت $(X_{b,d(1)}, X_{b,d(2)}, \dots, X_{b,d(N_b)})$ نشان دهیم. رابطه‌های (۹) و (۱۰) نحوه محاسبه مقادیر قله و دره طیفی را نشان می‌دهند

$$peak_b = \log\left(\frac{1}{\alpha N_b} \sum_{i=1}^{\alpha N_b} x_{b,d(i)}\right) \quad (9)$$

$$valley_b = \log\left(\frac{1}{\alpha N_b} \sum_{i=1}^{\alpha N_b} x_{b,d(N_b-i+1)}\right) \quad (10)$$

در این رابطه‌ها، α نشان دهنده ضریب همسایگی است که مقداری بین (۰/۵) خواهد بود. در این مقاله ما از مقدار ۰/۲ برای α استفاده کرده‌ایم. حال با توجه به مقدار به دست آمده مقدار کنتراست طیفی را با استفاده از رابطه (۱۱) محاسبه می‌کنیم.

کاهنده در طرفین (مانند همینگ) ضرب شده و طیف فوریه آن قاب را محاسبه می‌کنیم. سپس بانک فیلتر توزیع شده بر اساس معیار مل را بر روی طیف فوریه اعمال می‌کنیم. در نهایت با محاسبه لگاریتم انرژی هر زیرباند و اعمال تبدیل کوسینوس برای آن‌ها، ضرائب کپسترال مبتنی بر معیار مل را به دست می‌آوریم.

۴-۴- شدت سیگنال مبتنی برای معیار اکتاو(OBSI)^۱

در این روش، یک بانک فیلتر همپوشان مثلثی را با شروع از یک مقدار فرکانس پایه مانند ۲۷ هرتز و توزیع مبتنی بر اکتاو ایجاد کرده و بر روی طیف فوریه قاب اعمال می‌کنیم [۱۱]. لگاریتم انرژی هر زیرباند را می‌توان به عنوان بردار ویژگی در نظر می‌گیریم. علاوه بر آن می‌توان با اعمال تبدیل کوسینوس بر روی این مقادیر ضرائب کپسترال مبتنی بر معیار اکتاو(OFCC)^۲ را نیز به دست آورد. با توجه به اینکه از نظر آماری، توزیع نمونه‌های عددی را می‌توان با استفاده از مقدار میانگین، واریانس و گشتاور^۳ های آن توصیف کرد، در کنار لگاریتم انرژی و ضرائب کپسترال، ۴ ویژگی آماری لگاریتم انرژی زیرباندها را نیز محاسبه کرده و به بردار ویژگی اضافه می‌کنیم. همچنین از تقسیم لگاریتم انرژی هر زیرباند به زیرباند قبلی ویژگی OBSIR^۴ به دست می‌آید.

۴-۵- کنتراست طیفی مبتنی بر معیار اکتاو(OSC)^۵

این روش استخراج ویژگی که برای سیگنال‌های موسیقی پیشنهاد شده است [۶] سعی بر این دارد که با استفاده از مقادیر قله و دره هر زیرباند مشخصات طیفی سیگنال را نشان دهد. در حالت کلی قله‌های طیفی نشان دهنده مولفه‌های هامونیک سیگنال و دره‌های طیفی نشان دهنده

¹ Octave Based Signal Intensities(OBSI)

² Octave Frequency Cepstral Coefficient

³ Statistical Moment

⁴ Octave Based Signal Intensities Ratio(OBSIR)

⁵ Octave-based Spectral Contrast (OSC)

$$SC_b = \text{peak}_b - \text{valley}_b \quad (11)$$

با توجه به مقادیر نشان داده شده در جدول ۱، به ازای هر قاب از سیگنال ۸ مقدار کنتراست طیفی محاسبه می‌کنیم.

۴-۶- MSFM^۱ و MSCM^۲

بر خلاف ویژگی‌هایی که تا کنون مطرح کردیم، این دو روش [۳] در رده روش‌های استخراج ویژگی میان‌مدت و بلندمدت قرار دارند. روش کار بدین صورت است که ابتدا سیگنال را بر اساس معیار اکتاو به تعدادی زیرباند تقسیم کرده و با استفاده از رابطه (۱۲) انرژی تک‌تک پنجره‌های تحلیلی را در هریک از زیرباندها محاسبه می‌کنیم. فرض کنید می‌خواهیم این دو بردار ویژگی را با استفاده از پنجره تحلیلی به دست آوریم. ($q=1,2,\dots,Q$)

$$E_q[i] = \sum_{f=L_i}^{H_i} |X[f]|^2 \quad (12)$$

در این رابطه i نشان دهنده شماره زیرباند اکتاو، L_i و H_i به ترتیب نشان دهنده کران پایین و کران بالای زیرباند و $X[f]$ نیز طیف سیگنال است. حال این دو ویژگی را برای هر کدام از زیرباندها با استفاده از رابطه‌های (۱۳) و (۱۴) محاسبه می‌کنیم (روابط را بدون ساده کردن آن‌ها نشان داده‌ایم).

$$MSFM[i] = \frac{\sqrt{\prod_{q=1}^{Q/2} A_q[i]}}{\sum_{q=1}^{Q/2} A_q[i]} \quad (12)$$

$$MSCM[i] = \frac{\max(A_q[i])^{Q/2}}{\sum_{q=1}^{Q/2} A_q[i]} \quad (13)$$

در رابطه‌های (۱۲) و (۱۳) مقدار $A_q[i]$ نشان دهنده طیفی - رابطه (۴) - می‌باشد که پس از اعمال تبدیل فوریه بر روی مقادیر $E_q[i]$ به دست می‌آید.

۴-۷- سایر روش‌ها

روش‌های دیگری همانند استخراج ریتم، پیچ‌های چندتایی موسیقی، ویژگی‌های MPEG-7، سایر ویژگی‌های طیفی که در اینجا بررسی نکردیم نیز وجود دارند که برای آشنایی با جزئیات آن‌ها می‌توان به [۹] مراجعه نمود.

۵- تجمیع ویژگی‌ها

می‌دانیم که ویژگی‌های میان‌مدت نیز با استفاده از ویژگی‌های کوتاه‌مدت به دست می‌آیند. سوالی که در اینجا مطرح می‌شود اینست که چگونه می‌توان با استفاده از ویژگی‌های کوتاه‌مدت، ویژگی‌های میان‌مدت سیگنال موسیقی را به دست آورد؟ بدین منظور روش‌های مختلفی پیشنهاد شده اند که در ادامه به رایج‌ترین آن‌ها اشاره می‌کنیم.

مقدار میانگین و واریانس: در این روش، مقدار میانگین و واریانس چند بردار ویژگی کوتاه‌مدت متوالی را به عنوان ویژگی میان‌مدت سیگنال موسیقی استفاده می‌کنند علاوه بر مقدار میانگین و واریانس می‌توان مامنت‌های مرتبه ۳ و بالاتر بردارهای ویژگی کوتاه‌مدت را به عنوان بردار ویژگی میان‌مدت سیگنال نیز اضافه کرد [۱۳].

مقدار میانگین و کوواریانس: در این روش با محاسبه مقدار میانگین و کوواریانس بردارهای ویژگی کوتاه‌مدت، بردار ویژگی میان‌مدت را به دست می‌آورند [۱۳].

ضرائب بانک فیلتر: در این روش ابتدا چند بردار ویژگی کوتاه‌مدت متوالی را انتخاب کرده و سپس طیف فوریه تک تک ویژگی‌های موجود در بردارهای ویژگی کوتاه‌مدت را

¹ Modulation Spectral Flatness Measure

² Modulation Spectral Crest Measure

محاسبه می‌کنیم [۱۴]. در ادامه طیف فوریه به دست آمده از هر ویژگی را به ۴ زیرباند تقسیم می‌کنیم:

- زیرباند اول شامل فرکانس صفر هرتز یا همان مقدار میانگین بردار ویژگی
- زیرباند دوم شامل فرکانس ۱ الی ۲ هرتز یا همان انرژی مادولاسیون ویژگی‌ها
- زیرباند سوم شامل فرکانس ۳-۱۵ هرتز
- زیرباند چهارم شامل فرکانس ۱۵ تا ماکزیمم فرکانس طیف.

با این حال می‌توان تعداد زیر باندها را افزایش داد که نتایج تجربی نشان داده است افزایش تعداد زیر باند می‌تواند موجب بهبود کارایی سیستم گردد [۱۵].

ویژگی‌های آماری و شکل طیف: در این روش پس از محاسبه طیف فوریه تک تک ویژگی‌ها، مامنت‌های مرتبه ۱ تا ۴ آن‌ها محاسبه می‌شوند. همچنین می‌توان از ویژگی‌های شکلی طیف همانند شیب، مرکز ثقل و موارد مشابه دیگر نیز استفاده کرد [۱۳].

ضرائب پیشگویی خطی: ضرائب پیشگویی خطی یکی از روش‌های شناخته شده در رگرسیون سری‌های زمانی می‌باشد [۱۵]. با استفاده از همین روش نیز می‌توان بردارهای ویژگی کوتاه‌مدت متوالی را به صورت ضرایب پیشگویی خطی مدلسازی کرد. با توجه به اینکه در اینجا به جای مقادیر عددی زمانی با بردارهای عددی زمانی سروکار داریم، از اینرو ضرایب پیشگویی خطی که در این جا به دست می‌آوریم به جای یک بردار، ماتریسی از ضرایب خواهد بود که هر سطر از این ماتریس نشان دهنده ضرائب پیشگویی خطی یکی از ویژگی‌ها می‌باشد. برای محاسبه مقدار این ضرائب نیز می‌توان از روش دوربین-لوینسون استفاده کرد.

۶- پیاده‌سازی و آزمایش‌ها

در این فصل نتایج حاصل از پیاده‌سازی یک سیستم شناسایی سبک موسیقی را نشان می‌دهیم. برای استخراج ویژگی، جعبه ابزارها و نرم افزارهای مختلفی همانند **MARYAS**، **JPEMToolbox**، **AudorityToolbox** و **Voicebox** وجود دارد. با این حال به دلیل اینکه این جعبه ابزارها برخی از روش‌های استخراج ویژگی معرفی شده در این مقاله را پیاده‌سازی نکرده‌اند، از اینرو همه آن‌ها را در محیط **MATLAB R2008a** پیاده‌سازی کردیم. برای آموزش مدل مخلوط گاوسی با توجه به اینکه جعبه ابزار متلب دارای مشکلاتی در آموزش مدل‌هایی با ابعاد بالا بود، از توابع پیاده‌سازی شده توسط [۷] استفاده کرده و برای رتبه‌بندی ویژگی‌ها نیز نرم افزار **وکا (Weka)**^۱ را به کار بردیم.

در این آزمایش‌ها سعی بر آن داریم که ۸ سبک موسیقی مختلف شامل سبک‌های **Celtic**، **Classic**، **Classic Piano**، **Jazz**، **Metal**، **Persian Classic**، **Relaxing** و **Dance** را شناسایی کنیم. بدین منظور فایل‌های موسیقی مربوط به این سبک‌های موسیقی را از آلبوم‌هایی با نوازنده‌های مختلف جمع‌آوری کرده و همانطور که در جدول ۲ مشاهده می‌کنید، آن‌ها را به دو دسته آموزشی و آزمایشی تقسیم کردیم. مدت زمان اجرای هر فایل موسیقی به طور متوسط ۳ دقیقه، نرخ نمونه‌برداری ۴۴۱۰۰ نمونه در ثانیه به شکل مونو و با پسوند wav می‌باشد.

جدول ۲: مجموعه داده‌های آموزشی و آزمایشی

سبک	آموزشی	آزمایشی
Celtic	۲۹ فایل	۲۱ فایل
Classic	۴۴ فایل	۳۳ فایل
Classic piano	۴۴ فایل	۳۸ فایل
Jazz	۳۱ فایل	۲۷ فایل
Metal	۲۵ فایل	۲۰ فایل
Persian Classic	۵۲ فایل	۴۴ فایل
Relaxing	۵۴ فایل	۳۸ فایل
Dance	۵۵ فایل	۳۹ فایل

¹ WEKA : <http://www.cs.waikato.ac.nz/ml/weka/>

برای استخراج ویژگی، ابتدا سیگنال موسیقی را به قالب‌های ۴۰ میلی‌ثانیه با همپوشانی ۲۰ میلی‌ثانیه‌ای تقسیم کرده و به ازای هر قالب به ترتیب ۱۸ ویژگی OSC^۱، ۸ ویژگی طیفی^۲، ۱۳ ویژگی MFCC^۳، ۷ ویژگی OBSI^۴، ۶ ویژگی OBSIR^۵، ۴ ویژگی بردارهای مرتبه یک تا چهار OBSI، ۵ ویژگی OFCC^۶ را استخراج کردیم که در این حالت بردار ویژگی کوتاه‌مدت استخراج شده شامل ۶۱ ویژگی خواهد بود. سپس طول پنجره بافت را ۱ ثانیه با همپوشانی ۰/۵ ثانیه انتخاب کرده و برای تجمیع ویژگی‌های کوتاه‌مدت از روش میانگین (با بالانویشت^۷ M در کنار نام ویژگی‌ها مشخص شده است) و انحراف معیار (با بالانویشت^۷ D در کنار نام ویژگی‌ها مشخص شده است) استفاده کردیم. با این کار به ازای هر پنجره بافت، ۱۲۲ ویژگی به دست می‌آید که از آن‌ها برای مدل‌سازی سبک‌های موسیقی استفاده کردیم.

۶-۱- انتخاب بردار ویژگی مناسب برای شناسایی سبک موسیقی

با توجه به بالا بودن ابعاد بردار ویژگی، قبل از انجام مدل‌سازی، اعمال یک الگوریتم انتخاب ویژگی اجتناب‌ناپذیر است. برای انتخاب بردار ویژگی با استفاده از روش‌های آماری، از نرم افزار وکا استفاده کردیم. بدین منظور بردارهای ویژگی میان‌مدت به دست آمده از همه سبک‌ها را وارد نرم افزار وکا کرده و با استفاده از الگوریتم ChiSquaredAttributeEval به رتبه‌بندی بردارهای ویژگی پرداختیم. این الگوریتم از معیار Chi-Squared [۱۶] برای رتبه‌بندی ویژگی‌ها استفاده می‌کند. جدول ۳ نتایج این رتبه‌بندی را نشان می‌دهد.

ویژگی‌هایی که رتبه بالاتری دارند، از قدرت جداکنندگی بیشتری نیز برخوردار هستند. به عنوان مثال ویژگی ۱۸ام در

بردار ویژگی دارای رتبه نخست می‌باشد. با توجه به توضیحاتی که در بخش ۶ دادیم، ویژگی ۱۸ام برابر با میانگین آخرین ضریب ویژگی OSC خواهد بود. همچنین ویژگی‌های قرار گرفته در رتبه ۲ تا ۶ نیز متعلق به میانگین همین ضرایب می‌باشد. به لحاظ نظری، استفاده از ویژگی‌هایی با رتبه بالاتر برای دسته‌بندی الگوها نسبت به زمانی که از ویژگی‌هایی با رتبه پایین‌تر برای همان منظور استفاده کنیم، از اولویت بیشتری برخوردار است. برای اطمینان از صحت رتبه‌بندی انجام شده، سه آزمایش مختلف انجام داده و نتایج آن را باهم مقایسه کردیم. در آزمایش اول، ویژگی‌های با رتبه ۹۳ تا ۱۲۲ را به عنوان بردار ویژگی انتخاب کرده و ماتریس سردرگمی^۸ را به دست آوریم. سپس همین کار را دوباره برای ویژگی‌های با رتبه ۴۶ تا ۷۵ و ۱ تا ۳۰ به طور جداگانه انجام داده و دو ماتریس سردرگمی دیگر نیز به دست آوردیم. برای مدل‌سازی نیز توسط شبکه عصبی پرسپترون با دولایه مخفی (لایه اول ۱۲ و لایه دوم ۱۷ نورون با توابع فعالیت سیگموئید دوقطبی و لایه خروجی ۸ نورون با توابع فعالیت خطی) استفاده کردیم. همچنین هنگام تعیین داده‌های آموزشی برای شبکه عصبی، ۶۰٪ داده‌ها را به عنوان داده آموزشی، ۲۰٪ داده‌ها به عنوان داده اعتبارسنجی شبکه و ۲۰٪ باقی‌مانده داده‌ها نیز به عنوان داده‌های آزمایشی حین آموزش شبکه انتخاب شدند. جدول ۴ نتایج به دست آمده از این آزمایش برای بردارهای ویژگی مختلف را نشان می‌دهد.

نتایج به دست آمده از جدول ۴ نشان می‌دهند که ویژگی‌های با رتبه بالاتر عملکرد بهتری نسبت به ویژگی‌های با رتبه پایین دارند با این حال نتایج به دست آمده از ویژگی‌های رتبه میانی و ابتدایی بسیار به هم نزدیک هستند. با توجه به این که اعمال الگوریتم‌های تکاملی با تابع شایستگی دقت دسته‌بندی نیاز به صرف زمان بسیار زیادی دارد، از اینرو سعی کردیم با چند بار آزمون و خطا تخمینی بردار ویژگی مناسب را به دست آوریم. به همین منظور آزمایش‌هایی با ترکیب ویژگی‌های از رتبه‌های ابتدایی و

^۸ Confusion Matrix

^۱ Octave-based Spectral Contrast

^۲ Spectral Feature (SF)

^۳ Mel Frequency Cepstral coefficient

^۴ Octave Based Signal Intensities

^۵ Octave Based Signal Intensities Ratio

^۶ Octave Frequency Cepstral Coefficient

^۷ Superscript

میانی جدول انجام دادیم و در پایان نتایج حاصل نشان داد که با استفاده از ویژگی‌های رتبه ۱ تا ۲۰ و ۵۶ تا ۶۵ خطای دسته‌بندی کمتر از ۱۰٪ می‌باشد. جدول ۵ ماتریس سردرگمی به دست آمده برای این بردار ویژگی را نشان می‌دهد.

جدول ۴: نتایج به دست آمده از دسته‌بندی با ویژگی‌هایی با رتبه‌های مختلف بر حسب درصد

رتبه ۱ الی	رتبه ۴۶ الی	رتبه ۹۳ الی	
۳۰	۷۵	۱۲۲	
۷۱	۹۵	۶۶	سبک
۸۱	۷۵	۷۵	Celtic
۹۷	۱۰۰	۷۶	Classic
۸۵	۷۷	۸۵	Classic piano
۷۵	۷۵	۸۰	Jazz
۹۷	۸۸	۷۲	Metal
۸۴	۹۲	۷۳	Persian Classic
۹۲	۸۴	۸۷	Relaxing
۸۷	۸۶	۷۷	Dance

نتایج به دست آمده از جدول ۶ و جدول ۵ حاکی از آن است که شبکه‌های عصبی مصنوعی (جدول ۵) نتایج قابل قبول‌تری را نسبت به سایر روش‌ها دارند. با این حال مدل مخلوط گاوسی نیز تقریباً همانند شبکه عصبی می‌تواند به دسته‌بندی سبک‌های موسیقی بپردازد. همچنین با اینکه در روش‌های مدل مخلوط گاوسی شرطی و دوتایی سعی کردیم به طور صریح از توالی رخداد بردارهای ویژگی نیز استفاده کنیم، با این حال در عمل می‌بینیم که استفاده از تجمع ویژگی نیز به خوبی می‌تواند این توالی را مدل کند. علاوه بر آن نتایج به دست آمده از این دو روش نشان دهنده ضعف مدلسازی با استفاده از روش‌های گسسته همانند دوتایی می‌باشد. نکته حائز اهمیت در این آزمایش اینست که همه روش‌های دسته‌بندی دقت پایینی در دسته‌بندی سبک Celtic داشته‌اند. علت این مسئله را نیز می‌توان در شباهت زیاد این سبک با سبک‌های Metal و Classic پیدا کرد. در ادامه سبک‌های Jazz, Classic Piano, Classic, Celtic, Persian Classic, Metal, Relaxing و Dance را به ترتیب با حروف A تا H نشان می‌دهیم.

جدول ۵: ماتریس سردرگمی به دست آمده از بردار ویژگی حاصل از ویژگی‌های رتبه ۱ تا ۲۰ و ۵۶ تا ۶۵

	A	B	C	D	E	F	G	H	ACC
A	۱۶	۰	۲	۱	۰	۰	۰	۲	٪۷۶
B	۰	۲۷	۱	۰	۰	۴	۱	۰	٪۸۱
C	۰	۰	۳۸	۰	۰	۰	۰	۰	٪۱۰۰
D	۰	۰	۰	۲۴	۰	۱	۰	۲	٪۸۸
E	۰	۰	۰	۰	۱۸	۰	۰	۲	٪۹۰
F	۱	۰	۰	۰	۰	۴۳	۰	۰	٪۹۷
G	۰	۰	۰	۱	۰	۱	۳۶	۰	٪۸۴
H	۰	۰	۰	۱	۰	۰	۰	۳۸	٪۹۷
دقت تصمیم‌گیری									٪۹۰

۳-۶- بررسی توانایی روش‌های استخراج ویژگی در شناسایی سبک موسیقی

در این آزمایش قصد داریم توانایی دسته‌بندی هر یک از

۶-۲- مدل سازی با استفاده از سایر روش‌ها

در این آزمایش از سایر روش‌های مدلسازی برای شناسایی سبک موسیقی استفاده کردیم. بدین منظور برای هر یک از سبک‌های نشان داده شده با برچسب A تا H به ترتیب تعداد مولفه‌های گاوسی را ۵، ۱۵، ۱۰، ۱۰، ۱۰، ۲۰ و ۱۵ مولفه در نظر گرفته و با استفاده از روش مدل مخلوط گاوسی به شناسایی ۸ سبک موسیقی پرداختیم (تعداد مولفه‌های گاوسی که در اینجا استفاده کرده‌ایم بصورت تجربی و با سعی و خطا بدست آمده‌اند). سپس برای مدلسازی از درخت تصمیم و روش بیز ساده نیز استفاده کردیم.

در کنار این روش‌ها سعی کردیم از ماشین بردار پشتیبان کرنل توابع پایه شعاعی نیز استفاده کنیم. در نهایت با در نظر گرفتن توالی رخداد بردارهای ویژگی و مدلسازی با استفاده مدل مخلوط گاوسی و دوتایی‌ها، به مدلسازی این مساله پرداختیم.

جدول ۳: نتایج رتبه‌بندی ویژگی‌های میان‌مدت

رتبه	ویژگی	رتبه	ویژگی	رتبه	ویژگی	رتبه	ویژگی	رتبه	ویژگی	رتبه	ویژگی
۱	^M OSC ₁₈	۲۲	^D SF ₁	۴۳	^M SF ₈	۶۴	^M SF ₇	۸۵	^D OBSI ₃	۱۰۶	^D MCFF ₆
۲	^M OSC ₁₂	۲۳	^M MFCC ₈	۴۴	^M OBSI ₅	۶۵	^M OBSIR ₂	۸۶	^M OBSIR ₃	۱۰۷	^M OBSI ₁₀
۳	^M OSC ₁₆	۲۴	^M OSC ₈	۴۵	^M OFCC ₄	۶۶	^M OBSIR ₆	۸۷	^D OSC ₁	۱۰۸	^D MFCC ₃
۴	^M OSC ₁₇	۲۵	^D SF ₈	۴۶	^D OSC ₁₈	۶۷	^D OSC ₁₄	۸۸	^M OSC ₁	۱۰۹	^D OBSI ₆
۵	^M OSC ₁₀	۲۶	^M MFCC ₆	۴۷	^D SF ₃	۶۸	^D OSC ₁₂	۸۹	^D OBSI ₅	۱۱۰	^D OSC ₃
۶	^M OSC ₄	۲۷	^M MFCC ₂	۴۸	^D OBSIR ₁	۶۹	^D MFCC ₁	۹۰	^D OFCC ₃	۱۱۱	^D MFCC ₇
۷	^M MFCC ₁	۲۸	^D SF ₄	۴۹	^M OBSI ₆	۷۰	^D OBSIR ₆	۹۱	^M MFCC ₁₃	۱۱۲	^D MFCC ₉
۸	^M OBSI ₈	۲۹	^D SF ₅	۵۰	^M OSC ₉	۷۱	^D OBSI ₄	۹۲	^D OSC ₁₁	۱۱۳	^D OBSIR ₅
۹	^M OFCC ₁	۳۰	^M OSC ₆	۵۱	^D OBSI ₂	۷۲	^D OFCC ₂	۹۳	^D MFCC ₄	۱۱۴	^M OBSIR ₄
۱۰	^M OSC ₁₄	۳۱	^M MFCC ₅	۵۲	^M MFCC ₉	۷۳	^D OSC ₁₅	۹۴	^M OFCC ₅	۱۱۵	^D MFCC ₈
۱۱	^M SF ₃	۳۲	^M MFCC ₁₀	۵۳	^M OSC ₁₁	۷۴	^D OSC ₂	۹۵	^M OBSI ₉	۱۱۶	^D OSC ₉
۱۲	^M OBSI ₃	۳۳	^M OBSI ₇	۵۴	^M OSC ₇	۷۵	^M OBSIR ₅	۹۶	^D OBSI ₁₀	۱۱۷	^D MFCC ₁₀
۱۳	^M SF ₄	۳۴	^D OSC ₁₇	۵۵	^D OBSI ₁	۷۶	^D OSC ₈	۹۷	^M OSC ₅	۱۱۸	^D MFCC ₁₁
۱۴	^M SF ₁	۳۵	^M OFCC ₂	۵۶	^M OSC ₁₃	۷۷	^D OSC ₄	۹۸	^D OSC ₁₃	۱۱۹	^D OSC ₅
۱۵	^M OBSI ₂	۳۶	^D SF ₇	۵۷	^D OSC ₁₆	۷۸	^D OFCC ₁	۹۹	^D OBSIR ₄	۱۲۰	^D MFCC ₁₃
۱۶	^M OBSI ₄	۳۷	^M SF ₆	۵۸	^D OFCC ₅	۷۹	^D OBSI ₈	۱۰۰	^D OBSI ₁₁	۱۲۱	^D MFCC ₁₂
۱۷	^M SF ₂	۳۸	^M MFCC ₄	۵۹	^M OSC ₁₅	۸۰	^D OBSIR ₃	۱۰۱	^D OBSIR ₂	۱۲۲	^D OSC ₇
۱۸	^M OSC ₂	۳۹	^M MFCC ₃	۶۰	^M OFCC ₃	۸۱	^D OBSI ₉	۱۰۲	^M OBSI ₁₁		
۱۹	^M SF ₅	۴۰	^D SF ₆	۶۱	^D OSC ₁₀	۸۲	^D OSC ₆	۱۰۳	^D MFCC ₆		
۲۰	^M OBSI ₁	۴۱	^M MFCC ₁₂	۶۲	^D OFCC ₄	۸۳	^M MFCC ₁₁	۱۰۴	^M OSC ₃		
۲۱	^D SF ₂	۴۲	^M MFCC ₇	۶۳	^M OBSIR ₁	۸۴	^D MFCC ₂	۱۰۵	^D OBSI ₇		

جدول ۶: نتایج به دست آمده از مدلسازی با سایر روش‌ها بر حسب

درصد

مدل مخلوط گاوسی	درخت تصمیم	مدل مخلوط گاوسی تشریحی	دسته بندی بیز ساده	ماشین بردار پشتیبان	دقت
celtic	۷۱	۶۱	۷۶	۵۲	۵۷
classic	۸۷	۷۲	۹۰	۸۴	۷۵
Classic piano	۱۰۰	۹۷	۹۷	۹۲	۱۰۰
Jazz	۸۵	۸۸	۸۸	۸۱	۸۸
Metal	۱۰۰	۸۵	۹۵	۷۵	۸۰
Persian Classic	۹۳	۹۵	۹۳	۸۶	۹۵
Relaxing	۸۴	۸۶	۸۹	۴۴	۷۶
Dance	۹۲	۹۷	۹۴	۷۴	۹۴
میانگین دقت دسته بندی	۸۹	۸۵	۹۰	۷۳	۸۳

صدا توسط آلت‌های موسیقی پیدا کرد. در آلت‌های موسیقی ناپیوسته همانند پیانو، سنتور، چنگ و هرکدام از نت‌های آن‌ها دارای محدوده فرکانسی مشخصی می‌باشند که معمولاً از قانون نرخ فرکانسی ۱ به ۲ (اکتاو)، ۴ به ۵ (سوم)، ۳ به ۴ (چهارم) و ۲ به ۳ (پنجم) تبعیت می‌کنند. از اینرو توزیع فیلتر بانک بر این اساس می‌تواند مفید باشد. روش شدت سیگنال باندهای اکتاو در حقیقت با توزیع فیلترها بر اساس معیار اکتاو بردار ویژگی مربوط به یک سیگنال را استخراج می‌کند.

۴-۶- مدلسازی با استفاده از ساختار سلسله مراتبی

در آزمایش‌های قبلی از روش دسته بندی مسطح برای شناسایی سبک‌های موسیقی استفاده کردیم. در این آزمایش قصد داریم همین مسئله را با استفاده از یک دسته بندی سلسله مراتبی مدلسازی نماییم. بدین منظور با توجه به نوع سبک موسیقی، ساختار سلسله مراتبی را به شیوه‌ای که در شکل ۴ نشان داده شده است، انجام داده و نتایج شکل ۵ را به دست آوردیم.

این آزمایش نشان می‌دهد که در مساله ما ساختار سلسله مراتبی نیز تقریباً از دقتی مشابه دقت ساختار مسطح برخوردار است. با این حال استفاده از این ساختار به خصوص زمانی که با تعداد بسیاری از سبک‌های موسیقی مواجه هستیم می‌تواند مفید باشد.

۵-۶- ترکیب دسته بندی کننده‌ها

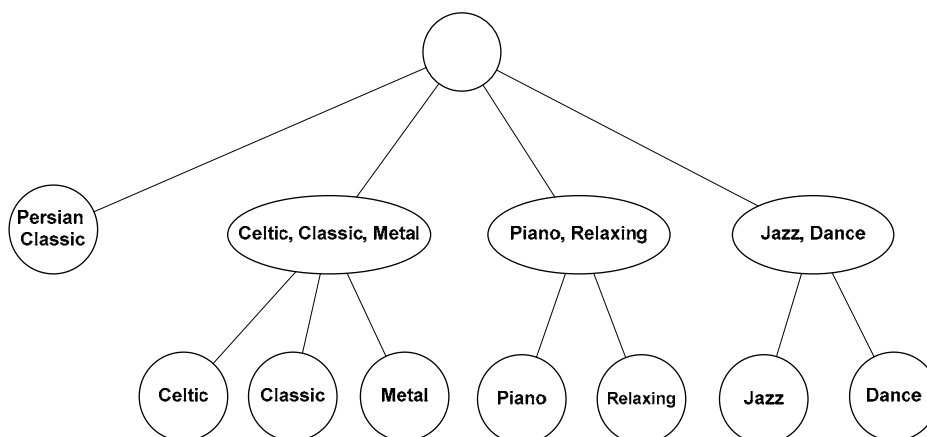
در این آزمایش ابتدا به طور جداگانه با استفاده از شبکه‌های عصبی مصنوعی، مدل مخلوط گاوسی و درخت تصمیم به

روش‌های استخراج ویژگی را به طور جداگانه مورد بررسی قرار دهیم. جدول ۷ نتایج به دست آمده از این آزمایش را نشان می‌دهند. در اینجا ما از شبکه عصبی به عنوان ابزار دسته بندی استفاده کرده ایم.

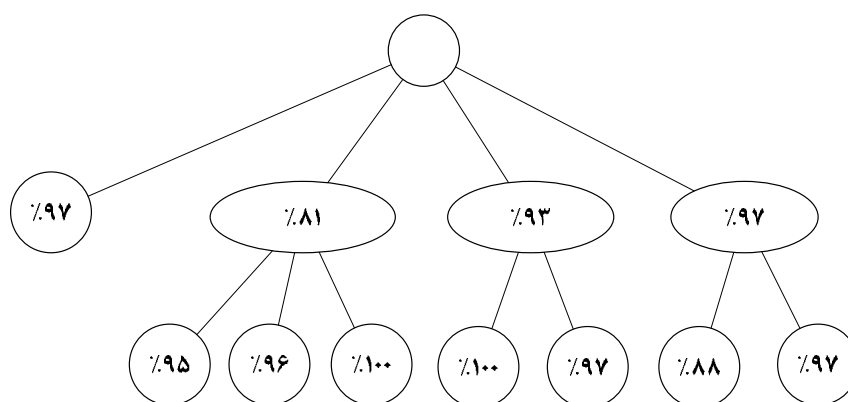
جدول ۷ نشان می‌دهد که کنتراست طیفی مبتنی بر معیار اکتاو از دقت بیشتری در دسته بندی سبک‌های موسیقی دارد. همانطور که در جدول ۳ نیز نشان دادیم، این ویژگی‌ها از نظر آماری نیز دارای رتبه بالاتری نسبت به سایر ویژگی‌ها هستند. روش استخراج ویژگی شدت و نسبت سیگنال باندهای اکتاو تقریباً مشابه روش ضرایب کپسترال مبتنی بر معیار مل می‌باشد. با این تفاوت که در آن توزیع فیلترها بر اساس معیار اکتاو می‌باشد. با اینکه نتایج جدول ۷ حاکی از دقت پایین این روش در مساله ما دارد، ولی در جدول ۳ می‌بینیم که سه مقدار از این بردار ویژگی دارای رتبه بالایی می‌باشند. علت این امر را نیز می‌توان در بازه فرکانسی تولید

جدول ۷: نتایج به دست آمده از روش‌های استخراج ویژگی مختلف بر حسب درصد

شدت و نسبت سیگنال باندهای اکتاو	ضرایب کپسترال مبتنی بر معیار مل	ویژگی‌های طیفی	کنتراست طیفی مبتنی بر معیار اکتاو	
۵۲	۶۶	۷۱	۶۶	Celtic
۷۲	۵۷	۸۴	۷۸	Classic
۵۵	۸۱	۸۱	۱۰۰	Classic piano
۷۰	۷۴	۵۱	۷۰	Jazz
۷۵	۹۵	۸۰	۸۵	Metal
۶۱	۹۷	۹۰	۹۳	Persian Classic
۶۰	۶۰	۵۵	۵۷	Relaxing
۷۹	۸۷	۸۷	۸۹	Dance
۶۵	۷۷	۷۵	۸۰	میانگین دقت دسته‌بندی



شکل ۴: ساختار دسته‌بندی سلسله مراتبی برای شناسایی سبک‌های موسیقی



شکل ۵: نتیجه دسته‌بندی به صورت سلسله مراتبی

مدلسازی مسئله پرداخته و سپس با ترکیب آن‌ها و استفاده از روش انتخاب مبتنی بر رای‌گیری به شناسایی سبک‌های موسیقی مختلف پرداختیم.

در این آزمایش از میان ۲۶۰ داده آزمایشی، تنها ۱۲ عدد آن‌ها دارای ابهام بوده و دسته‌بندی کننده‌ها برای تعیین کلاس آن‌ها به توافق نرسیدند و در بقیه موارد حداقل ۲ دسته‌بندی کننده بر روی سبک موسیقی شناسایی شده بر داده آزمایشی، توافق داشتند. این آزمایش نشان می‌دهد که برای افزایش دقت دسته‌بندی می‌بایست از روش‌های استخراج ویژگی جدیدی استفاده کرد. چراکه دسته‌بندی کننده‌ها تقریباً به شکل یکسانی به مدلسازی این مساله می‌پردازند.

حروف **A تا J** بکار گرفته شده است. نتایج به دست آمده از این تحقیق در جدول ۹ نشان داده شده است.

جدول ۹: نتایج به دست آمده در تحقیق [۱]

	A	B	C	D	E	F	G	H	I	J	%
A	۶۹	۰	۰	۰	۱	۰	۰	۰	۰	۰	۹۸
B	۰	۵۳	۲	۰	۵	۸	۶	۴	۲	۰	۶۶
C	۰	۸	۵۲	۱۱	۰	۱۳	۱۴	۵	۹	۶	۴۴
D	۰	۳	۱۶	۶۴	۱	۶	۳	۲۶	۷	۶	۴۸
E	۲۶	۴	۰	۰	۷۵	۸	۷	۱	۲	۱	۶۰
F	۵	۱۳	۴	۱	۹	۴۰	۱۴	۱	۷	۳۳	۳۱
G	۰	۷	۰	۱	۳	۴	۴۳	۱	۰	۰	۷۲
H	۰	۹	۱۰	۱۸	۲	۱۲	۱۱	۵۹	۷	۱	۴۵
I	۰	۲	۱۴	۵	۳	۵	۰	۳	۶۶	۰	۶۷
J	۰	۱	۰	۱	۰	۴	۲	۰	۰	۵۳	۸۶
دقت دسته‌بندی											۶۱

در [۴] نیز از روش مدلسازی سوم (رجوع به بخش ۲) به همراه ۱۲ ضریب کپسترال مبتنی بر معیار مل استخراج شده از پنجره‌های تحلیلی برای شناسایی سبک‌های موسیقی **Rock/Pop, Metal, Jazz/Blues, Electronic, Classic** و **World** (به ترتیب حروف **A تا F**) استفاده شده است. جدول ۱۰ نتایج به دست آمده از این تحقیق را نشان می‌دهد.

جدول ۱۰: نتایج به دست آمده در تحقیق [۴]

	A	B	C	D	E	F	ACC
A	۳۰۰	۱	۰	۰	۰	۱۹	۹۳
B	۲	۹۵	۱	۱	۷	۸	۸۳
C	۰	۳	۱۴	۰	۶	۳	۵۳
D	۰	۰	۰	۲۰	۲۳	۲	۴۴
E	۲	۱۵	۰	۴	۷۶	۵	۷۴
F	۱۲	۱۶	۰	۰	۱۲	۸۲	۶۷
دقت دسته‌بندی							۶۹

در [۶] از ضرایب کپسترال مبتنی بر معیار مل و کنتراست طیفی مبتنی بر معیار اکتاو (رجوع به بخش ۴) و ابزار مدلسازی نزدیکترین همسایه برای شناسایی ۷ سبک موسیقی **Jazz, HipHop, Dance, Chamber, Rock, Pop, Orchestra** (به ترتیب حروف **A تا G**)

۶-۷- مقایسه با نتایج سایر تحقیقات

بر خلاف مسئله شناسایی آلت موسیقی، در زمان نگارش این مقاله پایگاه داده استاندارد برای ارزیابی و مقایسه روش‌های شناسایی سبک موسیقی ارائه نشده است. از اینرو معمولاً محققین در مطالعاتشان پایگاه داده خاص خود را ایجاد کرده و ارزیابی روش خود را بر روی آن پایگاه داده انجام می‌دهند. در این بخش برخی از نتایج به دست آمده از سایر تحقیقات را مورد بررسی قرار می‌دهیم.

در [۱] از میانگین و واریانس (رجوع به بخش ۵) ویژگی‌های زمانی و طیفی معرفی شده در بخش ۴ (۸ ویژگی)، پنج ضریب اول ضرایب کپسترال مبتنی بر معیار مل (۱۰ ویژگی) ، کمترین انرژی (۱ ویژگی)، ویژگی‌های ریتمی (۶ ویژگی) و ویژگی‌های گام (۵ ویژگی) به عنوان بردار ویژگی (۳۰ بعدی) استخراج شده از کل سیگنال (روش مدلسازی اول معرفی شده در بخش ۲) استفاده شده است. در نهایت از این بردار ویژگی به همراه ابزارهای مدل مخلوط گاوسی، دسته‌بندی کننده گاوسی و **k** نزدیکترین همسایه برای شناسایی ۱۰ سبک موسیقی **HipHop, Disco, Country, Classic, HipHop, Metal, Pop, Regge, Blues, Rock, Jazz** (به ترتیب

استفاده شده است. نتایج این تحقیق را در جدول ۱۰ قابل مشاهده است. بر خلاف سایر جدول‌ها، مقادیر نشان داده شده در هر یک از درایه‌های جدول ۱۱ نشان دهنده درصد دسته‌بندی می‌باشد.

جدول ۱۱: نتایج به دست آمده در تحقیق [۶]

	A(%)	B(%)	C(%)	D(%)	E(%)	F(%)	G(%)
A(%)	۷۷/۷۸	۲۱	۰	۰	۳/۵	۰	۰
B(%)	۱۶/۸۵	۸۰/۹	۰	۰	۲/۲۵	۰	۰
C(%)	۰	۰	۸۹/۶	۶/۴۴	۰/۵	۱/۹۸	۱/۴۹
D(%)	۰	۰	۱۲/۱	۸۲/۴۲	۰	۲/۲	۳/۳
E(%)	۱۰/۸۹	۳/۹۶	۰	۲/۹۷	۶۸/۳۲	۹/۹	۳/۹۶
F(%)	۱	۱	۱	۳	۱۴	۷۲	۸
G(%)	۰	۰	۱/۴۸	۳/۷	۱/۴۸	۱۲/۵۹	۸۰/۷۴

همانطور که مشاهده می‌کنید نتایج به دست آمده در سایر تحقیقات به دلیل آنکه تنها بر روی روش استخراج ویژگی خاصی تمرکز کرده‌اند در مقایسه با نتایج به دست آمده در تحقیق ما از دقت بسیار کمتری برخوردار هستند. این در حالی است که ما ابتدا ویژگی‌ها مختلفی را از داده‌های آموزشی استخراج کرده و سپس با استفاده از یک روش رتبه‌بندی ویژگی سعی کردیم بردار ویژگی با قدرت تفکیک‌پذیری بالا را انتخاب کنیم. با دقت در جدول ۱۱ و نتایج نشان داده شده در جدول ۳ می‌توان به این نتیجه رسید که روش استخراج ویژگی OSC^۱ در مقایسه با سایر روش‌ها، عملکرد بهتری را برای حل مساله شناسایی سبک موسیقی از خود نشان می‌دهد.

۷- جمع‌بندی

شناسایی خودکار سبک موسیقی مساله‌ای است که امروزه مطالعات گسترده‌ای بر روی آن در حال انجام است. با این حال به دلیل تنوع زیاد سبک‌های موسیقی و همچنین شباهت آن‌ها به همدیگر، روش‌های پیشنهاد شده برای آن از دقت بسیار بالایی برخوردار نیستند.

توجه به این نکته حائز اهمیت است که انسان نیز در شناسایی سبک‌های موسیقی دارای دقت بسیار بالایی نمی‌باشد. روش‌های مدل‌سازی مختلفی برای حل این مساله وجود دارد که در حالت کلی به دو دسته تقسیم می‌شوند. در روش‌های اول توالی رخداد بردارهای ویژگی به‌طور صریح مدنظر ما نبوده و با تجمیع ویژگی‌ها به شکل ضمنی این توالی را نشان می‌دهیم. ولی در روش‌های مدل‌سازی دوم به طور صریح این توالی زمانی را مورد توجه قرار داده و بر اساس آن سبک‌های موسیقی را مدل می‌کنیم.

مسئله دیگر در شناسایی موسیقی، ویژگی‌هایی است که برای سیگنال موسیقی باید استخراج کنیم. روش‌های استخراج ویژگی را از نظر حوزه استخراج می‌توان به ۳ دسته کلی تقسیم کرد: (۱) ویژگی‌های زمانی همانند نرخ عبور از صفر، انرژی و ضرایب خودهمبستگی که در حوزه زمان استخراج می‌شوند. (۲) ویژگی‌های طیفی همانند وضوح و شدت طیفی، پیچ و کنتراست طیفی مبتنی بر معیار اکتاو سیگنال که در حوزه فرکانس استخراج می‌شوند. (۳) ویژگی‌های کپسترال همانند ضرایب کپسترال مبتنی بر معیار مل یا ضرایب کپسترال خطی که اعمال تبدیل کوسینوس بر روی خروجی فیلتر بانک اعمال شده در حوزه فرکانس به دست می‌آیند. در یک دسته‌بندی دیگر ویژگی‌های استخراج شده را از نظر زمانی می‌توان به سه دسته (۱) کوتاه مدت (۲۰-۴۰ میلی‌ثانیه) (۲) میان مدت (۱ الی ۳ ثانیه) و (۳) بلند مدت تقسیم کرد. ویژگی‌های میان مدت معمولاً با استفاده از ویژگی‌های کوتاه مدت محاسبه می‌شوند، که این عمل را در اصطلاح تجمیع ویژگی می‌گویند. روش‌های تجمیع ویژگی که بر روی چند بردار ویژگی کوتاه مدت اعمال می‌شوند نیز به سه دسته عمده تقسیم می‌گردند: (۱) روش‌های آماری همانند محاسبه میانگین، واریانس و **گشتاور**های تک تک ویژگی‌ها (۲) روش‌های طیفی مانند ضرایب بانک فیلتر که با محاسبه طیف هر یک از ویژگی‌ها به دست می‌آیند و (۳) ماتریس ضرایب پیشگویی خطی. با ترکیب روش‌های مدل‌سازی، استخراج ویژگی و همچنین تجمیع ویژگی می‌توانیم

¹ Octave-based Spectral Contrast

سیستم‌هایی با قابلیت‌های متفاوت به دست می‌آوریم. نکته مهمی که در سال‌های اخیر بیشتر مورد توجه محققین قرار گرفته است استفاده از روش دسته‌بندی سلسله مراتبی به جای روش مسطح است که می‌تواند دقت سیستم شناسایی سبک موسیقی را افزایش داده و آن را برای شناسایی تعداد زیادی از سبک‌های موسیقی قادر سازد.

۸- نتیجه‌گیری

مسئله شناسایی سبک موسیقی به دلیل اینکه همپوشانی زیادی بین سبک‌های موسیقی مختلف وجود دارد، یکی از زمینه‌های فعال در عرصه پردازش سیگنال می‌باشد. به عنوان مثال سبک **Jazz** و **Celtic** هم از نظر ریتمیک مشابه هم بوده و از نظر صدای آلت‌های موسیقی نیز شباهت‌های بسیاری میان این دو سبک وجود دارد. همچنین برخی از آهنگ‌های موسیقی ترکیبی از سبک‌های مختلف می‌باشد.

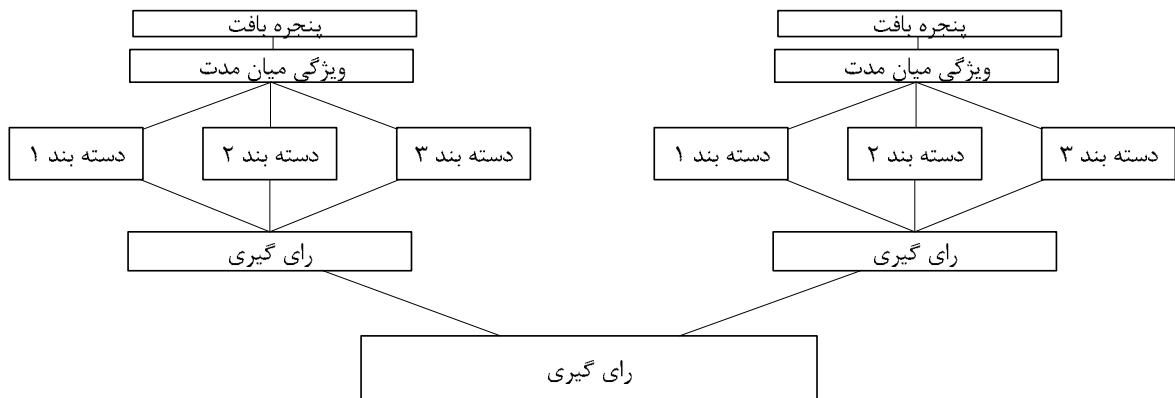
به عنوان مثال آهنگ **Secret Garden** ابتدا با سبک موسیقی **Classic Piano** شروع شده و در ادامه تبدیل به سبک موسیقی **Classic** می‌شود. این در حالی است که سهم هرکدام از این سبک‌ها در شکل‌گیری این آهنگ یکسان می‌باشد. بنابراین در برخی آهنگ‌های موسیقی با یک مفهوم غیرقطعی در خصوص سبک موسیقی مواجه هستیم. با توجه به اینکه شناسایی سبک موسیقی در بیشتر موارد با استفاده از روش رای‌گیری انجام می‌گیرد، بنابراین توسط هرکدام از روش‌های دسته‌بندی می‌توان میزان دخیل بودن هر یک از سبک‌های موسیقی در یک آهنگ را مشخص نمود.

نتایج به دست آمده از رتبه‌بندی آماری و دسته‌بندی سبک‌های موسیقی نشان می‌دهد که ضرایب کنتراست طیفی مبتنی بر معیار اکتاو از قدرت جداکنندگی بالایی برخوردار هستند. همچنین ویژگی‌های طیفی نیز از نظر آماری دارای رتبه بالایی هستند. برای به دست آوردن ویژگی‌های میان مدت سیگنال، روش میانگین و واریانس رایج بوده و نتایج قابل قبولی را از خود نشان می‌دهد. با

توجه به اینکه این روش تجمیع ویژگی به طور ضمنی می‌تواند توالی رخداد بردارهای ویژگی را مدل کند، از اینرو استفاده از روش‌های مدل‌سازی همانند دوتایی‌ها که به طور صریح این توالی را مدل می‌کنند، مفید نخواهد بود. از میان روش‌های مختلف دسته‌بندی نیز مدل مخلوط گاوسی و شبکه‌های عصبی نتایج رضایت بخشی را برای دسته‌بندی سبک‌های موسیقی از خود نشان داده‌اند.

همچنین نشان دادیم که استفاده از ساختار سلسله مراتبی در مساله‌ای که ما تعریف کردیم، عملکرد مشابه با روش دسته‌بندی مسطح از خود نشان می‌دهد. با این حال استفاده از ساختار سلسله مراتبی برای تعیین سبک موسیقی در مواردی که با تعداد زیادی از سبک‌های موسیقی مواجه هستیم، مفید خواهد بود. علاوه بر آن، با استفاده از روش رای‌گیری و ترکیب سه دسته‌بندی کننده سعی کردیم دقت دسته‌بندی را افزایش دهیم. اما نتایج به دست آمده نشان دادند که حتی ترکیب دسته‌بندی کننده‌ها موجب بهبود عملکرد سیستم شناسایی سبک موسیقی نمی‌شود. بلکه به منظور کاهش خطای دسته‌بندی، نیاز به روش استخراج ویژگی جدیدی داریم. بر اساس مطالعات ما و آزمایش‌های صورت گرفته، این روش استخراج ویژگی باید بتواند ریتم و الگوی زیرومی موسیقی را نشان دهد. به عنوان مثال سبک‌های موسیقی همانند **Classic** و **Rap** کاملاً دارای ریتمی مجزا می‌باشند. همچنین سبک‌های موسیقی **Celtic** و **Classic** ریتم تقریباً مشابهی دارند و یکی از خصوصیت‌هایی که این دو سبک را از هم جدا می‌کند، الگوی زیرومی آن‌ها است. با اینکه روش‌های استخراج ویژگی بیان شده در این تحقیق نیز تا حدودی این دو خصوصیت را از سیگنال موسیقی استخراج می‌کنند ولی با این حال نیازمند مطالعات بیشتری در این زمینه است.

از جمله نکاتی که تا کنون در هیچیک از روش‌های استخراج ویژگی مدنظر قرار نگرفته است، الگوی نواختن آلت‌های موسیقی در یک سیگنال موسیقی است. در اینجا دو مسئله مطرح می‌شود. مسئله اول شناسایی نوع آلت موسیقی بوده و مسئله دوم الگوی استفاده از آلت‌های موسیقی در یک سبک



شکل ۶: نحوه شناسایی سبک‌های موسیقی با ترکیب دسته‌بندی کننده‌های مختلف

جدول ۸: نتایج به دست آمده از ترکیب دسته‌بندی کننده‌های مختلف

	A	B	C	D	E	F	G	H	Acc.
A		۰	۳	۱	۰	۰	۰	۲	%۷۱
B	۰		۳	۰	۰	۲	۱	۰	%۸۱
C	۰	۰		۰	۰	۰	۰	۰	%۱۰۰
D	۰	۰	۰		۰	۱	۰	۲	%۸۸
E	۲	۰	۰	۰		۰	۰	۱	%۸۰
F	۲	۰	۰	۰	۰		۰	۰	%۹۵
G	۰	۰	۰	۱	۰	۱		۰	%۹۴
H	۰	۰	۰	۰	۰	۱	۰		%۹۷
								دقت دسته‌بندی	%۸۸
A: Celtic, B: Classic, C: Classic Piano, D: Jazz, E: Metal, F: Persian Classic, G: Relaxing, H: Dance									

می‌توان آن‌ها را از هم تفکیک نمود. از سوی دیگر آلت موسیقی همانند فولوت و ویولون در سبک موسیقی Classic به ترتیب به عنوان آلت موسیقی پس‌زمینه و اصلی استفاده می‌شود. این در حالی است که در سبک موسیقی Relaxation همان آلت‌های موسیقی به ترتیب به عنوان آلت موسیقی اصلی و پس‌زمینه بکار می‌روند. به‌طور خلاصه می‌توان چنین بیان نمود در مساله شناسایی سبک موسیقی تفکیک آلت‌های موسیقی، تشخیص پس‌زمینه و زمینه بودن

موسیقی خاص است. بنابراین الگوی نواختن موسیقی بدین معنا است که ابتدا باید بتوانیم در یک سیگنال موسیقی که ممکن است ده‌ها آلت موسیقی به طور همزمان نواخته می‌شود، هریک از آن‌ها را از هم جدا کرده و سپس الگوی نواختن آن‌ها را در هریک از سبک‌های موسیقی به طور مجزا مدل کنیم. به عنوان مثال آلت موسیقی ویولون به طور عمده در سبک موسیقی Persian و New Class، Classic، Classic نواخته می‌شود. از سوی دیگر تنها با گوش دادن به الگوی نواختن ویولون در هریک از این سه سبک به راحتی

[10] Tero Tolonen, Matti Karjalainen, A Computationally Efficient Multipitch Analysis Model, IEEE Transactions on Speech and Audio Processing, 2000, vol. 8 Issue:6, pp. 708-716.

[11] Slim Essid, Gaël Richard, Bertrand David, Musical Instrument Recognition by Pairwise Classification Strategies, IEEE Transactions on Audio, Speech, and Language Processing, 2006, vol. 14 Issue:4, pp. 1401-1412.

[12] Tao Li, Mitsunori Ogihara, Qi Li, A Comparative Study on Content-Based Music Genre Classification, international ACM SIGIR conference on Research and development in information retrieval, 2003, pp. 282-289.

[13] Cyril Joder, Slim Essid, Gaël Richard, Temporal Integration for Audio Classification with Application to Musical Instrument Classification, IEEE Transactions on Audio, Speech, and Language Processing, 2009, vol. 17 Issue:1, pp. 174-186.

[14] Martin F. McKinney, Jeroen Breebaart, Features for Audio and Music Classification, International conference on music information retrieval, 2003, pp. 151-158.

[15] Anders Meng, Peter Ahrendt, Jan Larsen, Improving Music Genre Classification by Short-Time Feature Integration, International Conference on Acoustics, Speech, and Signal Processing, 2005, Vol. 5, pp.497-500.

[16] Huiqing Liu, Jinyan Li, Limsoon Wong 2002, A Comparative Study on Feature Selection and Classification Methods Using Gene Expression Profiles and Proteomic Patterns, Journal of Genome Informatics, vol.13, pp.51-60.

[17] Nicolas Scaringella, Giorgio Zoia, and Daniel Mlynek, Automatic Genre Classification of Music Content: A Survey, Signal Processing Magazine, 2006, vol.23 Issue:2, pp. 133 – 141.

[18] Stefan Brecheisen, Hans-Peter Kriegel, Peter Kunath, Alexey Pryakhin, Hierarchical Genre Classification for Large Music Collections, International Conference on Multimedia and Expo, 2006, pp. 1385 – 1388.

[19] Qian Ding, Nian Zhang, Classification of Recorded Musical Instruments Sounds Based on Neural Networks, Symposium on Computational Intelligence in Image and Signal Processing, 2007, pp. 157-162.

هریک آلتها و همچنین الگوی نواختن تک تک آنها یک زمینه تحقیقاتی فعال در این عرصه می باشد.

مراجع

[1] Tzanetakis George, Cook Perry, Musical Genre Classification of Audio Signals, IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, 2002, vol.10 Issue:5, pp.293-302.

[2] Silla Carlos N., Kaestner Celso A. A., Koerich Alessandro L., Automatic Music Genre Classification Using Ensemble of Classifiers, IEEE International Conference on Systems, Man and Cybernetics, 2007, pp.1687-1692.

[3] Jang Dalwon, Jin Minho, Yoo Chang D., Music Genre Classification Using Novel Features and a Weighted Voting Method, International Conference on Multimedia and Expo, 2008, pp.1377-1380.

[4] Langlois Thibault, Marques Gonc, Automatic Music Genre Classification Using a Hierarchical Clustering and a Language Model Approach, International Conference on Advances in Multimedia, 2009, pp. 188-193.

[5] Lee Jae-Won, Park Soo-Beom, Kim Sang-Kyoon, Music Genre Classification Using a Time-Delay Neural Network, Lecture Notes in Computer Science, 2006, vol. 3972/2006, pp. 178-187.

[6] Lee Chang-Hsing, Shih Jau-Ling, Yu Kun-Ming, Su Jung-Mau, Automatic Music Genre Classification Using Modulation Spectral Contrast Feature, IEEE International Conference on Multimedia and Expo, 2007, pp. 204-207.

[7] Calinon S., Guenter F., Billard A., On Learning, Representing and Generalizing a Task in a Humanoid Robot, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2007, vol. 37 Issue:2, pp. 286-298.

[8] Li Tao, Ogihara Mitsunori, Music Genre Classification with Taxonomy, International Conference on Acoustics, Speech, and Signal Processing, 2005, Vol. 5, pp.197-200.

[9] Hyoung-Gook Kim, Nicolas Moreau, Thomas Sikora, Audio Classification Based on MPEG-7 Spectral Basis Representations, IEEE Transactions on Circuits and Systems for Video Technology, 2004, vol. 14 Issue:5, pp. 716-725.

Automatic Recognition of Music Genre

Hamed Habib Aghdam

Mohammad Mehdi Homayounpour

Laboratory for Intelligent Processing of Signal and Speech, Amirkabir University of Technology, Tehran, Iran.

Email: {habibiaghdam, homayoun}@aut.ac.ir

Computer Engineering Department, Amirkabir University of Technology

{habibiaghdam, homayoun}@aut.ac.ir

Abstract

Nowadays, automatic analysis of music signals has gained a considerable importance due to the growing amount of music data found on the Web. Music genre classification is one of the interesting research areas in music information retrieval systems. In this paper several techniques were implemented and evaluated for music genre classification including feature extraction, feature selection and music genre modeling on a database of 8 different music genres containing Celtic, Classic, Classic Piano, Jazz, Metal, Persian Classic, Relaxing and Dance music. This database was gathered from several albums composed by different musicians. Short, middle and long term features were studied and finally only short and middle term features were used in our experiments. The long term features were discarded due to their low performance in music genre classification. Two modeling types of the music genres were evaluated. In the first type, only distribution of the feature vectors was used and in the second type, the ordering of the feature vectors was taken into account. Some modeling techniques such as ANN, GMM, Decision Tree and SVM were used individually and in a hierarchical approach. We proposed a taxonomy which classifies the music genres in a hierarchy where there are a small number of classes in the root and large number of classes in leaves. In fact, each class at the root of taxonomy contains one or more music genres and each genre is represented as a leaf at the bottom of the taxonomy. In addition, several classifiers were used simultaneously, in a way that each of them classifies the music genres individually. The decision is finally made using a voting algorithm. Besides, several short-term feature extraction techniques which have successfully been applied in speech recognition, music instrument classification and also music genre classification were studied and after analysis of the experimental results using statistical measures and different combinations of features, a near optimal feature vector was selected.

Keywords: Automatic recognition of music genre, spectral feature, octave base spectral contrast, octave based signal intensities, mel frequency cepstral coefficients, feature integration