بازیابی موسیقی مبتنی بر نمونه با کاربرد تشخیص ژانر جهت بهبود سرعت عملکرد

نسترن برجیان*[1]، احسان‌اله کبیر[1]، ساناز سیدین[2]، الیپس مسیحی[3]

۱. دانشکده مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس

۲. دانشکده مهندسی برق، دانشگاه صنعتی امیرکبیر

۳. دانشکده فنی و مهندسی، دانشگاه تربیت مدرس

چکیده

هدف یک سامانه بازیابی اطلاعات موسیقی با دریافت یک نمونه، بازیابی آهنگ متناظر با نمونهٔ پرس‌وجوی کاربر از یک دادگان خاص است. این نمونه می‌تواند یک قطعهٔ چند ثانیه‌ای ضبط‌شده از هر منبع پخش‌کننده موسیقی مانند تلویزیون یا حتی یک محیط نوفه‌ای، مانند ورزشگاه باشد. در این مقاله، یک سامانه بازیابی اطلاعات موسیقی مبتنی بر نمونه با کاربرد تشخیص ژانر پیشنهاد شده است، که هدف آن، نشان‌دادن اثر کاربرد تشخیص ژانر موسیقی برای دستیابی به عملکرد دقیق و سریع در چنین سامانه‌هایی حتی در حضور نوفهٔ پس‌زمینه است. این سامانه شامل دو بلوک تشخیص ژانر و تطبیق- بازیابی است. در تشخیص ژانر از یک درخت تصمیم دودویی و در تطبیق- بازیابی از دو فاصلهٔ اقلیدسی و واگرایی کول‌بک- لیبلر (کِی‌اِل) به‌همراه یک روش ترکیب تصمیم مبتنی بر امتیازدهی استفاده شده است. سامانه پیشنهادی بر روی دادگان شناخته‌شدهٔ جی‌تی‌زان (ارائه شده توسط جرج زانتاکیس) و با دو دستهٔ تصادفی از علامت‌های پرس‌وجوی خالص و نوفه‌ای ارزیابی شد. نتایج، دستیابی به صحت ۹۷٪ و ۸۶٪ را به ترتیب برای دو حالت خالص و نوفه‌ای در زمان بازیابی ۵۲۵ میلی‌ثانیه با فاصلهٔ اقلیدسی نشان می‌دهند. این مقادیر برای فاصلهٔ کِی‌اِل به ۹۷٪ و ۸۲٪ با زمان بازیابی ۳۸۰ میلی‌ثانیه می‌رسند.

کلیدواژه‌ها: بازیابی اطلاعات موسیقی، پرس‌وجو با دریافت نمونه، تشخیص ژانر، ترکیب تصمیم، نوفه.

۱. مقدمه

امروزه، تولید محتوای صوتی و تصویری با سرعت فزاینده‌ای در حال رشد می‌باشد، و در هر ثانیه، تعداد بیش‌تری از این محتواهای رسانه‌ای تولید می‌شوند. هم‌زمان با آن، پیشرفت چشم‌گیر در فن‌اوری ذخیره‌سازی رقومی[1]، امکان ذخیره‌سازی هزاران نمونه اطلاعات مانند متن، تصویر، ویدئو و صدا را در یک حافظه کوچک فراهم ساخته است [۱-۲]. بنابراین، ایجاد امکان جستجوی دقیق در بین این حجم گستردهٔ اطلاعات جهت بازیابی اطلاعات مورد نظر اجتناب‌ناپذیر می‌باشد [۳-۵]. در واقع، هم‌گام با پیشرفت‌های قابل توجه در سامانه‌های بازیابی تصویر در طی دههٔ اخیر [۶-۷]، سامانه‌های بازیابی صوتی مخصوصاً در زمینهٔ موسیقی نیز توسعهٔ چشم‌گیری یافته‌اند [۸].

در روندنمای یک سامانه بازیابی اطلاعات موسیقی[2] نوعی [۹-۱۱]، تعدادی ویژگی موسیقیایی از دادگان[3] استخراج و و ذخیره می‌شوند. اولین گام در فرآیند بازیابی، استخراج ویژگی‌های همسان از علامت پرس‌وجوی کاربر[4] است. سامانه بازیابی موسیقی، این ویژگی‌ها را با ویژگی‌های دادگان مقایسه کرده و تعدادی از موسیقی‌های نزدیک‌تر به علامت پرس‌وجو را بازیابی و به کاربر اعلام می‌کند [۱۲-۱۵].

با توجه به ماهیت سامانه‌های بازیابی موسیقی و تنوع آن‌ها و هم‌چنین، با در نظر گرفتن نیازهای کاربران، این سامانه‌ها را می‌توان از نقطه نظرهای مختلفی دسته‌بندی کرد. به‌عنوان مثال، در یک دسته‌بندی می‌توان سامانه‌های بازیابی موسیقی را از نظر نوع علامت پرس‌وجو، روش تطبیق[5] و نوع دادگان بررسی کرد [۱۲، ۱۶].

معمول‌ترین نوع‌های علامت پرس‌وجو در سامانه‌های

*نویسنده پاسخگو: nastaran.borjian@modares.ir

[1] Digital

[2] Music information retrieval

[3] Dataset

[4] Query

[5] Matching method

بازیابی موسیقی عبارتند از: یک نمونه[1] از موسیقی مطلوب [۵، ۱۷-۱۹] آواز[2] [۲۰-۲۱]، زمزمه[3] [۲۱-۲۲] و سوت[4]. علاوه بر تنوع در علامت پرس‌وجو، روش‌های جستجو و تطبیق مختلفی نیز در این سامانه‌ها به‌کار رفته است. به‌عنوان مثال، در منبع [۲۳] یک روش اِن-گرامز[5] برای جستجو و تطبیق استفاده شده است. در منبع [۲۴] مشخصات موسیقیایی نت‌ها[6] و در منابع [۲۵] و [۲۶]، استخراج ملودی[7] به‌عنوان پایهٔ مقایسه و تطبیق به‌کار رفته است. در منبع [۵] تابع چگالی احتمال بردارهای ویژگی دادگان و علامت پرس‌وجو با یکدیگر مقایسه می‌شوند. انواع دادگان موسیقی به‌کاررفته در سامانه‌های بازیابی موسیقی را نیز می‌توان به دو دستهٔ نوع تک‌آوایی[8] و نوع چندآوایی[9] [۲۷] تقسیم کرد. در دسته‌بندی دیگر می‌توان انواع دادگان را به دو گروه میدی[10] [۲۳] و هنجار[11] تقسیم‌بندی کرد.

علاوه بر موارد فوق، خود ماهیت موسیقی نیز از تنوع زیادی برخوردار است؛ و موسیقی‌ها در زمان‌ها، مکان‌ها و موقعیت‌های اجتماعی مختلف نواخته و اجرا شده‌اند. این تنوع در ماهیت موسیقی به صورت ژانر[12] مطرح می‌شود. سامانه‌های مختلفی جهت طبقه‌بندی ژانر در سال‌های اخیر طراحی و پیاده‌سازی شده‌اند که از لحاظ ویژگی‌های مورد استفاده، تعیین میزان شباهت، تعریف نوع علامت پرس‌وجو و معیار جداسازی دارای گوناگونی زیادی هستند [۴، ۲۸-۳۰].

از دیدگاهی دیگر، هم‌گام با پیشرفت‌های شایان در ابزارها و سامانه‌های رسانه‌ای در چند دههٔ گذشته، طراحی و پیاده‌سازی سامانه‌های بازیابی موسیقی در قالب وب‌سایت و یا برنامه‌های کاربردی روی گوشی‌های همراه نیز از پیشرفت قابل توجهی برخوردار بوده‌اند [۳۱-۳۳]. موتورهای جستجوگر فری‌دی‌بی[13] [۳۴] و میوزیک‌برینز[14] [۳۳]، از جملهٔ این سامانه‌های بازیابی موسیقی در قالب وب‌سایت هستند، که مبتنی بر دریافت یک نمونه علامت پرس‌وجو می‌باشند و به صورت برخط در دسترس کاربران قرار دارند.

از دیگر سامانه‌های بازیابی موسیقی مبتنی بر دریافت یک نمونه علامت پرس‌وجو که به صورت نرم‌افزار و برنامهٔ کاربردی روی گوشی‌های همراه ارائه شده‌اند، می‌توان به سوندهوند[15] [۳۵] و نیورُس[16] [۳۶] اشاره کرد. برنامه کاربردی شازام[17] [۳۱] نیز از جمله مشهورترین سامانه‌های بازیابی موسیقی روی گوشی‌های همراه است که هم در برابر نسخه‌های متعدد یک آلبوم و هم در برابر نوفهٔ پس‌زمینه[18] مقاوم است، و می‌تواند با درصد موفقیت بالایی نمونه علامت پرس‌وجوی دریافتی را از یک محیط با نوفهٔ پس‌زمینه، مانند یک ورزشگاه شناسایی کند. این سامانه از روش‌های اثر انگشت صدا برای جستجوی آهنگ مطلوب کاربر که در حال پخش از طریق رادیو یا در یک محیط تفریحی- ورزشی بوده و کاربر قطعه‌ای از آن را ضبط کرده، استفاده می‌کند. در یک آزمایش خاص که در منبع [۳۷] با هدف بررسی مقاومت سامانه شازام در مقابل نوفهٔ پس‌زمینه انجام شده است، نوفهٔ سفید گوسی[19] با نسبت‌های مختلف علامت به نوفه (اِس‌اِن‌آر)[20]، به علامت‌های پرس‌وجو افزوده شده، و میزان صحت بازیابی حاصل از سامانه شازام، در این حالات اندازه‌گیری شده است.

نرم‌افزار کاربردی شازام به عنوان یک سامانه مبتنی بر نمونه از روش شناسایی بیشینه‌های طیف بسآمد- زمان برای ایجاد یک مجموعه ویژگی تنک در یک بازهٔ بسآمدی خاص استفاده کرده، و علامت‌های پرس‌وجو با طول زمانی تا ۱۵ ثانیه را حتی هنگامی که از طریق گوشی همراه و از یک محیط نوفه‌ای مانند ورزشگاه دریافت شده‌اند، بازیابی می‌کند. از دیگر سامانه‌های بازیابی موسیقی مبتنی بر نمونه می‌توان به کار انجام شده در منبع [۳۸] اشاره کرد

[1] Example
[2] Singing
[3] Humming
[4] Whistling
[5] N-grams
[6] Notes
[7] Melody
[8] Monophonic
[9] Polyphonic
[10] MIDI; Musical Instrument Digital Interface
[11] Normal
[12] Genre
[13] Free DB

[14] MusicBrainz
[15] SoundHound
[16] Neuros
[17] Shazam
[18] Background noise
[19] Gaussian white noise
[20] SNR; Signal to Noise Ratio

[DOR: 20.1001.1.23455748.1395.4.1.2.9]

که جهت بازیابی نسخه‌های مختلف[1] آهنگ‌ها، معرفی شده شده است. این سامانه، آهنگ‌های هدف را براساس ملودی مشابه جستجو کرده و آهنگ‌هایی که دارای ملودی پایهٔ مشابه با علامت پرس‌وجو هستند را به عنوان خروجی برمی‌گرداند؛ در حالی‌که، ممکن است آهنگ با یک زبان متفاوت اجرا شده باشد یا حتی به‌وسیله خواننده‌های مختلف خوانده شده باشد. در منبع [۵] نیز یک سامانهٔ طبقه‌بندی سمعی بر مبنای نمونه[2] پیشنهاد شده است، که که از روش پارامتری کردن علامت به‌وسیله الگوی ترکیب گوسی (جی‌اِم‌اِم)[3] و به‌کارگیری این پارامترها برای جستجو در دادگان با استفاده از معیارهای شباهت مختلف، استفاده می‌کند. در منبع [۳۹] یک سامانهٔ بازیابی موسیقی مبتنی بر نمونه، با کاربرد یک الگوریتم جداسازی منبع صدا پیشنهاد شده است. در این سامانه، سه گروه از علامت‌های صوتی مبتنی بر سازهای دِرام[4] و گیتار و هم‌چنین صدای انسان از علامت پرس‌وجو جداسازی شده، و به‌وسیله یک واحد مهار[5] و توازنِ شدت صوتی پردازش می‌شوند. سپس یک مرحلهٔ بازترکیب که معادل با جابه‌جایی ژانر موسیقیایی است، بر روی علامت‌های پرس‌وجو، با هدفِ یافتن نتایج بازیابی از چند ژانر مشخص انجام می‌شود. ژانرهای مد نظر در این سامانه شامل کلاسیک[6]، دانس[7]، جاز[8] و راک[9] هستند.

بازیابی موسیقی با دریافت یک نمونه، مخصوصاً در مواقعی که کاربر بخشی از یک موسیقی مطلوب را ضبط کرده و در اختیار دارد، ولی هیچ اطلاعی مانند نام موسیقی، نام آلبوم و نام خواننده را از آن ندارد و تمایل دارد آن را جستجو و بازیابی کند، مورد اقبال کاربران است. علاوه بر این، از آن‌جا که نمونه علامت پرس‌وجو ممکن است در یک محیط نوفه‌ای مانند رستوران، ورزشگاه، یا گردهمایی، یا از تلویزیون، و یا حتی از یک دستگاه ضبط و پخش قدیمی ضبط شود و کاربر بخواهد آن را بازیابی کند، لذا در این حالت، وجود نوفهٔ پس‌زمینه که به‌صورت یک صدای ناخوشایند در ذیل آهنگ اصلی به گوش می‌رسد، اجتناب‌ناپذیر است. در چنین شرایطی نیز معیار مطلوبیت سامانه از نظر کاربر، قابلیت بازیابی آهنگ مورد نظر است. در این مقاله، یک سامانه بازیابی اطلاعات موسیقی با دریافت یک نمونه علامت پرس‌وجو مبتنی بر کاربرد تشخیص ژانر پیشنهاد می‌شود. در این سامانه با استفاده از نوفه سفید گوسی، اثر افزایش نوفه به‌صورت کلی مورد ارزیابی قرار می‌گیرد. دورنمای سامانه بازیابی موسیقی پیشنهاد شده، قابل استفاده بودن به عنوان یک نرم‌افزار کاربردی بر روی گوشی‌همراه و یا یک وب‌سایت بازیابی موسیقی است. برای چنین سامانه‌ای، صحت بازیابی و سرعت بازیابی دو معیار مقبولیت می‌باشند و سامانه پیشنهادی باید بتواند در کوتاهترین زمان ممکن و با صحت بالا، علامت پرس‌وجوی کاربر را حتی در شرایط وجود نوفهٔ پس‌زمینه بازیابی کند. در این مقاله، سعی شده است اثر کاربرد تشخیص ژانر در چنین سامانه‌ای برای حصول معیارهای موفقیت فوق بررسی شود.

در واقع، تشخیص ژانر موسیقی و بازیابی موسیقی به عنوان دو مقولهٔ پژوهشی مجزا در حوزهٔ موسیقی مورد توجه قرار می‌گیرند. اما در این مقاله پیشنهاد شده که از مقولهٔ تشخیص ژانر در راستای هدف اصلی (بازیابی موسیقی) استفاده شود. مشخصهٔ ژانر به‌خوبی منعکس‌کنندهٔ تفاوت ماهوی بین نوع‌های مختلف موسیقی است، و لذا تفکیک موسیقی‌ها با ویژگی‌های گوناگون را تسهیل می‌کند. علاوه بر این، در سامانه پیشنهادی در این مقاله، تشخیص ژانر به صورت خودکار انجام شده و هیچ لزومی به دانستن یا انتخاب ژانر علامت پرس‌وجو توسط کاربر نیست.

سامانه پیشنهادی در این مقاله شامل دو بلوک اصلی است که عبارتند از: تشخیص ژانر و تطبیق- بازیابی. در هر دو بلوک از ضرایب کپسترال مِل (اِم‌اِف‌سی‌سی)[10] به عنوان ویژگی استفاده شده است. ضرایب کپسترال مِل به‌خوبی منعکس‌کنندهٔ ویژگی‌های بافتی و طنینی[11] موسیقی هستند. برای تشخیص ژانر، استفاده از یک درخت تصمیم دودویی[12] پیشنهاد شده و در مرحله تطبیق- بازیابی از دو

[1] Cover versions
[2] Query-by-example audio classification system
[3] GMM; Gaussian Mixture Model
[4] Drama
[5] Control
[6] Classic
[7] Dance
[8] Jazz
[9] Rock

[10] MFCC; Mel-Frequency Cepstral Coefficients
[11] Timbral and textural features
[12] Binary decision tree

[DOR: 20.1001.1.23455748.1395.4.1.2.9]

فاصلهٔ اقلیدسی و واگرایی[1] (کِی‌اِل)[2] به‌همراه یک روش ترکیب تصمیم مبتنی بر امتیازدهی، جهت بازیابی آهنگ متناظر با علامت پرس‌وجو بهره گرفته شده است.

سامانه پیشنهادی بازیابی موسیقی بر روی دادگان شناخته‌شدهٔ جی‌تی‌زان (ارائه شده توسط جرج زانتاکیس)[3] زانتاکیس)[3] [۴۰] و با استفاده از دو دسته علامت پرس‌وجوی تصادفی به نام «علامت پرس‌وجوی خالص» و «علامت پرس‌وجوی نوفه‌ای» ارزیابی شده است. دادگان فوق، چندآوایی و بدون هیچ محدودیت بر روی تعداد نت‌های هم‌زمان می‌باشند. هم‌چنین، دادگان فوق به صورت ترکیب آوای خواننده و صدای سازهای موسیقی بوده، که کاملاً تداعی موسیقی معمول و رایج بین کاربران است. علامت پرس‌وجو نیز به صورت یک قطعهٔ چند ثانیه‌ای از موسیقی چندآوایی در نظر گرفته شده است. هر کدام از دو دسته علامت پرس‌وجوی خالص و نوفه‌ای مشتمل بر صد نمونهٔ تصادفی می‌باشد که دستهٔ نوفه‌ای با افزودن مقادیر مختلف نوفه سفید گوسی به دستهٔ خالص به‌دست آمده است.

نتایج سامانه بازیابی موسیقی پیشنهادی با دریافت یک نمونه علامت پرس‌وجو، دستیابی به صحت و زمان بازیابی مناسب را برای دو دسته علامت پرس‌وجوی خالص و نوفه‌ای نشان می‌دهد. تحقق این نتایج با استفاده از یک مرحله تشخیص ژانر مناسب جهت کاهش هزینه‌های محاسباتی و زمان بازیابی به همراه پیشنهاد یک مرحله تطبیق- بازیابی سریع و دقیق با استفاده از فنّ ترکیب تصمیم میسر شده است. علاوه بر این، جهت نشان دادن اثر تشخیص ژانر در کاهش زمان بازیابی، سامانه بازیابی فوق در حالت جستجوی پایه و بدون استفاده از این مرحله نیز مورد ارزیابی قرار گرفته است. هم‌چنین اثر تعدادی ویژگی شناخته شده زمانی- بسآمدی همراه با ویژگی‌های کپسترال مِل نیز در بلوک تطبیق- بازیابی مورد بررسی قرار گرفته‌اند. هدف از این بررسی، نشان دادن اثر انتخاب بهینه ویژگی‌ها در یک روش بازیابی موفق در مقابل استفاده از مجموعه‌ای متنوع از ویژگی‌ها با افزودن هزینه‌های غیرضروری محاسباتی است.

در بخش دوم مقاله، کلیت سامانه پیشنهادی بازیابی موسیقی به همراه مرحلهٔ تشخیص ژانر و مرحلهٔ تطبیق- بازیابی توضیح داده شده است. در بخش سوم، نتایج پیاده‌سازی تشریح و نهایتاً در بخش چهارم، نتیجه‌گیری ارائه گردیده‌اند.

۲. کلیت سامانه پیشنهادی بازیابی موسیقی

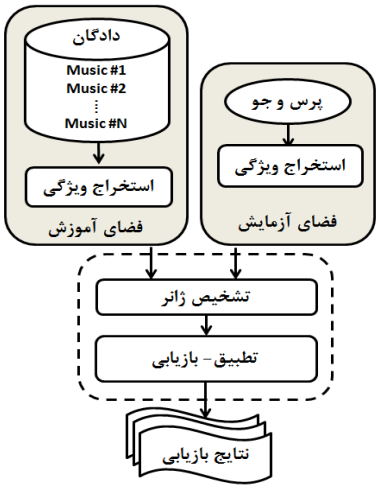
در این مقاله، یک سامانه بازیابی موسیقی با دریافت یک نمونه علامت پرس‌وجو مبتنی بر تشخیص ژانر پیشنهاد شده و اثر افزایش نوفه بر عملکرد سامانه با استفاده از نوفه سفید گوسی به صورت کلی مورد بررسی قرار گرفته است. در سامانه پیشنهادی، از تعریف ارائه شده در منبع [۱۶] برای این‌گونه سامانه‌های بازیابی موسیقی استفاده شده است. در واقع این تعریف، مبنای طراحی نرم‌افزارهای کاربردی بازیابی موسیقی بر روی گوشی‌های‌همراه می‌باشد. در این روش، علامت پرس‌وجو، یک نمونهٔ موسیقی چند ثانیه‌ای است و هدف، بازیابی آهنگ متناظر آن از پایگاه دادگان است. در این روش، پیشنهاد شده از تشخیص ژانر برای کاهش فضای جستجو و هزینه‌های محاسباتی استفاده شود.

در این مقاله از دادگان جی‌تی‌زان استفاده شده است [۴۰]. این دادگان شامل ۱۰۰۰ موسیقی در ۱۰ ژانر متفاوت می‌باشند. هر ژانر شامل ۱۰۰ عدد موسیقی با طول ۳۰ ثانیه و بسآمد نمونه‌برداری ۲۲۰۵۰ هرتز است. ژانرهای دادگان عبارتند از: بلوز[4]، کلاسیک، کانتری[5]، دیسکو[6]، هیپ هاپ[7]، جاز، متال[8]، پاپ[9]، رگا[10] و راک. با با توجه به نمونه سامانه‌های بازیابی موسیقی ذکر شده در مقدمه و با توجه به طول داده‌های استفاده شده در این مقاله که ۳۰ ثانیه هستند؛ طول علامت پرس‌وجو برابر ۵ ثانیه انتخاب شد. از نظر موسیقیایی نیز، ۵ ثانیه حدوداً مدت زمانی است که کفایت می‌کند تا شنونده‌ای ژانر و نام قطعه موسیقی پخش شده را از میان یک پایگاه دادگان معلوم تشخیص دهد.

[1] Divergence
[2] KL; Kullback-Leibler
[3] GTZAN; George Tzanetakis

[4] Blues
[5] Country
[6] Disco
[7] Hip hop
[8] Metal
[9] Pop
[10] Regae

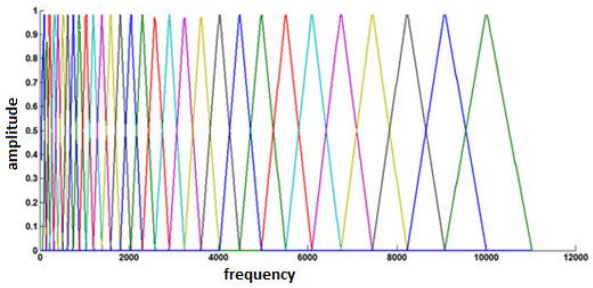
[DOR: 20.1001.1.23455748.1395.4.1.2.9]



شکل ۱ بلوک دیاگرام سامانه پیشنهادی بازیابی موسیقی؛ بلوک تشخیص ژانر و بلوک تطبیق- بازیابی، بلوک‌های اصلی در این سامانه هستند. ژانر علامت پرس‌وجو در بلوک تشخیص ژانر مشخص شده و پس از آن در بلوک تطبیق- بازیابی، آهنگ‌های ژانر مربوطه با استفاده از معیارهای فاصلهٔ اقلیدسی و کِی‌اِل سنجیده شده، و در نهایت آهنگ متناظر علامت پرس‌وجو شناسایی و بازیابی می‌شود.

در این سامانه، علامت پرس‌وجو به صورت تصادفی از دادگان انتخاب می‌شود. بنابراین همه ژانرها و همه داده‌های موسیقی موجود در یک ژانر، دارای شانس یکسان جهت انتخاب علامت پرس‌وجو هستند. علاوه بر این، علامت پرس‌وجو از هر نقطه‌ای از دادهٔ موسیقی متناظر می‌تواند شروع شود. به عبارت دیگر، نقطه شروع علامت پرس‌وجو نیز تصادفی است. بلوک دیاگرام سامانه بازیابی موسیقی پیشنهادی در شکل ۱ نشان داده شده است. بلوک‌های اصلی آن عبارتند از: استخراج ویژگی، تشخیص ژانر و تطبیق- بازیابی.

ابتدا داده‌های موسیقی دادگان، قالب‌بندی می‌شوند. طول قالب‌ها، ۴۰ میلی‌ثانیه بوده و با یکدیگر ۵۰ درصد هم‌پوشانی دارند. معمولاً در پژوهش‌های مرتبط با علامت‌های موسیقی، طول قالب برابر ۲۰ تا ۶۰ میلی‌ثانیه در نظر گرفته می‌شود تا بتوان علامت را در این بازه زمانی به‌صورت شبه ایستا فرض کرد و از روش‌های پردازش اطلاعات مبتنی بر فرض ایستا بودن علامت در این بازه‌ها استفاده کرد. فرایند قالب‌بندی و استخراج ویژگی‌های همسان بر روی علامت پرس‌وجو نیز انجام می‌شود. البته از آن‌جا که نقطه شروع علامت پرس‌وجو به صورت تصادفی انتخاب می‌شود، لذا هیچ الزامی در مورد انطباق قالب‌های علامت پرس‌وجو با قالب‌های داده موسیقی متناظر وجود ندارد.



شکل ۲ بانک صافی جهت محاسبهٔ ضرایب کپسترال مِل. این بانک شامل ۳۰ عدد صافی است که این صافی‌ها در بازه بسآمدی صفر تا ۱۱۰۲۵ هرتز تعریف شده‌اند. همان‌گونه که در شکل مشخص است بازهٔ غیر صفر این صافی‌ها به صورت لگاریتمی تغییر می‌کند. بخش غیر صفر صافی‌ها به شکل مثلث و با حداکثر دامنهٔ یک است. با این بانک صافی، انرژی علامت به صورت لگاریتمی در بازه‌های مختلف بسآمدی مقیاس‌بندی شده و متناظر آن، از هر صافی یک عدد ضریب کپسترال مِل حاصل می‌شود.

برای استخراج ضرایب کپسترال مِل، یک بانک صافی[1] مطابق شکل ۲ تعریف شده است. این بانک، شامل ۳۰ عدد صافی در بازه بسآمدی صفر تا ۱۱۰۲۵ هرتز است و از هر قالب، ۳۰ عدد ضریب کپسترال مِل با استفاده از این بانک صافی استخراج می‌شود. این ضرایب، بردار ویژگی را برای هر قالب تشکیل می‌دهند. انتخاب تعداد ضرایب کپسترال مِل و در پی آن، انتخاب تعداد صافی‌ها در بانک صافی براساس نتایج تجربی که در بخش نتایج شرح داده خواهد شد، صورت گرفته‌اند.

اگرچه تاکنون ویژگی‌های طنینی و بافتی زیادی برای پردازش علامت‌های موسیقی به‌کار رفته است؛ ولی در سامانه پیشنهادی بازیابی موسیقی از ضرایب کپسترال مِل به عنوان معمول‌ترین ویژگی‌های طنینی و بافتی استفاده شده است [۴۱-۴۴]. زیرا هدف، انتخاب بهترین ویژگی‌های ممکن جهت کسب نتایج مطلوب و بدون تحمیل هیچ‌گونه هزینهٔ اضافهٔ محاسباتی بوده است.

در مرحله تشخیص ژانر، یک درخت تصمیم دودویی پیاده‌سازی شده است، که با استفاده از کل بردارهای ویژگی دادگان و برچسب‌های متناظر آن‌ها آموزش داده می‌شود. این طبقه‌بند[2]، وظیفه تشخیص ژانر را برای

[1] Filter
[2] Classifier

علامت پرس‌وجوی کاربر بر عهده دارد. به عبارت دیگر، علامت پرس‌وجو، نقش دادهٔ آزمایشی را برای طبقه‌بند ایفا می‌کند.

در مرحله تطبیق- بازیابی از دو فاصلهٔ اقلیدسی و کِی‌اِل به‌همراه یک روش ترکیب تصمیم مبتنی بر امتیازدهی جهت بازیابی آهنگ متناظر با علامت پرس‌وجو بهره گرفته شده است. این مرحله با استفاده از یک پنجره لغزان به‌همراه محاسبهٔ فاصله و تخصیص امتیاز در هر لغزش، مطابق شکل ۳ انجام می‌شود. در شکل ۴، روندنمای روش ترکیب تصمیم پیاده‌سازی شده است. در این مرحله نیز از ضرایب کپسترال مِل استفاده می‌شود. البته علاوه بر ضرایب کپسترال مِل، تعدادی ویژگی شناخته‌شدهٔ زمانی- بسآمدی نیز در این مرحله مورد ارزیابی قرار گرفته است.

در یک سامانه بازیابی موسیقی با دریافت یک نمونه علامت پرس‌وجو، فقط کافی است که کاربر یک وسیلهٔ ضبط‌کنندهٔ صدا مانند گوشی‌همراه را که در حال پخش موسیقی مورد نظر کاربر می‌باشد، نزدیک یک منبع صوتی گرفته و چند ثانیه از این موسیقی را ضبط کند. سامانه بازیابی موسیقی باید توانایی بازیابی موسیقی مورد نظر را البته به شرط وجود در دادگان متناظر داشته باشد. بنابراین یک کاربر ممکن است علامت پرس‌وجو را در یک محیط نوفه‌ای مانند رستوران، ورزشگاه و یا محل‌های عمومی ضبط کرده باشد. در حالتی دیگر، ممکن است علامت پرس‌وجو از تلویزیون، یک دستگاه ضبط و پخش قدیمی یا از نوار کاست ضبط شده باشد. در همهٔ این حالات وجود نوفهٔ پس‌زمینه در علامت پرس‌وجو، که منجر به کیفیت پایین آن می‌گردد، اجتناب‌ناپذیر است.

قطعاً انواع نوفه‌های پس‌زمینهٔ فوق را می‌توان با انواع توابع نوفه (منطبق با محیط‌های خاص) الگوسازی کرد. با توجه به نمونه آزمایش انجام‌شده بر روی سامانه شازام [۳۷] و هم‌چنین نمونه کارهای بیان شده در [۴۵] و [۱۴] که از نوفه سفید گوسی جهت بررسی اثر کلی نوفه بر عملکرد سامانه استفاده شده است، در این مقاله نیز، عملکرد سامانه پیشنهادی با استفاده از نوفه سفید گوسی به صورت کلی ارزیابی می‌شود. البته با آزمون شنیداری تجربی که به کمک متخصص موسیقی انجام گرفت؛ این نتیجه‌گیری کلی به‌دست آمد که در مواردی که نوفه پس‌زمینه به صورت یک صدای یکنواخت ناخوشایند در ذیل آهنگ مطلوب به گوش می‌رسد، مثلاً زمانی که یک سامانه خنک‌کننده در محیط ضبط آهنگ مشغول به کار بوده یا آهنگ از یک دستگاه ضبط و پخش قدیمی ضبط شده است، می‌توان نوفه پس‌زمینه را تقریباً به عنوان نوفه ایستا پذیرفت، و درنظرگرفتن فرض شبه گوسی و الگو کردن آن با نوفه سفید گوسی می‌تواند تا حدی قابل قبول باشد.

در آزمون شنیداری تجربی که زیر نظر متخصص موسیقی انجام شد، مقادیر مختلف نوفه سفید گوسی به آهنگ‌ها افزوده شد و این آهنگ‌ها شنیده شدند. نتیجهٔ این آزمون تجربی نشان داد در مواردی که می‌توان نوفهٔ پس‌زمینه را در دنیای واقعی به‌صورت شبه گوسی فرض کرد، معمولاً مقدار علامت به نوفه بیش از ۴۰ دسی‌بل نیست.

جهت ارزیابی سامانه بازیابی موسیقی پیشنهادی در این مقاله، دو دسته علامت پرس‌وجوی خالص و نوفه‌ای آماده‌شد، که هر دسته شامل صد نمونهٔ تصادفی علامت پرس‌وجو می‌باشد. دستهٔ علامت پرس‌وجوی نوفه‌ای با افزودن مقادیر مختلف نوفه سفید گوسی به دستهٔ علامت پرس‌وجوی خالص ایجاد شده‌اند.

۲-۱. تشخیص ژانر

یکی از مقوله‌های معمول در زمینهٔ بازیابی اطلاعات موسیقی که در طی چند دههٔ اخیر به‌طورگسترده‌ای مورد توجه قرار گرفته است، تشخیص ژانر یک نمونهٔ موسیقی می‌باشد [۱۶]. در تشخیص ژانر، دادگان همراه با برچسب‌های ژانر متناظر به‌عنوان داده‌های آموزشی در فضای ویژگی به چندین طبقه الگوسازی می‌شوند. سپس از طریق یادگیری پارامترهای این طبقه‌های برچسب‌شده توسط یک طبقه‌بند مناسب، امکان تخصیص یک طبقه و متعاقباً یک برچسب ژانر، به یک علامت پرس‌وجوی ناآشنا میسر می‌شود.

در این مقاله، یک درخت تصمیم دودویی مبتنی بر الگوریتم کارت[1] با معیار انشعاب جی‌دی‌آی[2] برای تشخیص ژانر پیاده‌سازی شده است [۴۶-۴۹]. این طبقه‌بند با استفاده از کل بردارهای ویژگی استخراج‌شده از

[1] CART; Classification and Regression Tree
[2] GDI; Gini Diversity Inde

[DOR: 20.1001.1.23455748.1395.4.1.2.9]

دادگان، آموزش داده می‌شود. یادآوری می‌شود که هر بردار ویژگی شامل ۳۰ عدد ضریب کپسترال مِل محاسبه‌شده به ازای هر قالب از دادگان است. علامت پرس‌وجو نقش دادهٔ آزمایشی را برای طبقه‌بند دارد. وظیفه درخت تصمیم، تشخیص طبقه هر قالب از علامت پرس‌وجو می‌باشد. در نهایت، با یک رأی‌گیری حداکثری، ژانری که عمدهٔ قالب‌های علامت پرس‌وجو متعلق به آن تشخیص داده‌شده است، به‌عنوان ژانر علامت پرس‌وجو در نظر گرفته می‌شود.

در یک الگوریتم تشخیص ژانر صرف، بخشی از داده‌های موسیقی برای آموزش طبقه‌بند استفاده شده و بخشی دیگر برای آزمایش آن لحاظ می‌شوند. بنابراین دو دستهٔ داده‌های آموزشی و آزمایشی هیچ هم‌پوشانی با هم ندارند. ولی در سامانه بازیابی موسیقی پیشنهادی در این مقاله، صرفاً از مقولهٔ تشخیص ژانر به‌عنوان یک عامل کمکی در راستای افزایش سرعت بازیابی بهره‌گرفته شده است. برای حصول این خواسته، کل داده‌های موسیقی دادگان برای آموزش طبقه‌بند استفاده شده‌اند. در مقابل، علامت پرس‌وجو، نقش دادهٔ آزمایشی را برای طبقه‌بند ایفا می‌کند. هم‌چنین در این روش، طبقه‌بند به صورت برون‌خط پیاده‌سازی شده و آموزش داده می‌شود. لذا تعداد داده‌های مورد استفاده برای آموزش آن و زمان صرف‌شده جهت هم‌گرایی و یادگیری آن، معیار اصلی در بازیابی نمی‌باشد. بلکه مدت زمان فراخوانی طبقه‌بند در حین بازیابی، معیار کلیدی بوده و رابطهٔ مستقیم با حجم طبقه‌بند و حافظه لازم جهت ذخیره‌سازی آن دارد. در سامانه پیشنهادشده در این مقاله، و به‌منظور ایجاد حالت عملیاتی واقعی، مدت زمان فراخوانی طبقه‌بند برای هر علامت پرس‌وجو به‌عنوان بخشی از زمان لازم جهت بازیابی آن علامت پرس‌وجو درنظرگرفته شده است.

۲-۱-۱. درخت تصمیم دودویی کارت

درخت تصمیم پیاده‌سازی‌شده در این مقاله، شامل تعدادی گره تصمیم، تعدادی گره برگ و یک عدد گره ریشه است. هر گره تصمیم، براساس مقادیر عددی یک ویژگی منشعب می‌شود. در درخت تصمیم دودویی، تنها دو انشعاب یا شاخه به ازای هر گره تصمیم وجود دارد. تولید انشعاب‌ها در گره تصمیم با استفاده از یک آستانه‌گذاری بر روی یکی از ویژگی‌ها انجام می‌شود، به‌گونه‌ای که، یک معیار انشعاب از پیش تعریف‌شده را به بهترین نحو ممکن محقق سازد. سپس داده‌های ارزیابی‌شده در آن گره به دو گروه بزرگ‌تر و کوچک‌تر از آن آستانه تقسیم می‌شوند. هر انشعاب‌سازی، یک لایه به درخت تصمیم اضافه می‌کند. گره ریشه در برگیرندهٔ یکی از ویژگی‌ها به‌همراه یک آستانهٔ خاص است، به‌گونه‌ای که این ویژگی به‌همراه آستانهٔ مربوطه، داده‌های آموزشی را در بهترین حالت به دو گروه تقسیم می‌کنند. گره‌های برگ در واقع گره‌های انتهایی درخت بوده و هر یک از آن‌ها دارای یکی از برچسب‌های طبقهٔ تعریف‌شده در داده‌های آموزشی هستند.

در الگوریتم کارت سعی می‌شود در هر گرهٔ تصمیم، بهترین انشعاب‌سازی برای داده‌های آموزشی متناظر آن گره انتخاب شود. هدف در انشعاب‌سازی در درخت تصمیم دودویی، ایجاد دو گره جدید از یک گره مولد است به‌گونه‌ای که داده‌های متناظر هر گره جدید نسبت به داده‌های گره مولد، از تنوع طبقه‌ای کم‌تری برخوردار باشند. در این الگوریتم، انشعاب‌سازی به صورت متوالی انجام شده تا زمانی که برای گره جدید امکان تحقق شرط فوق وجود نداشته و لذا تداوم انشعاب‌سازی ممکن نباشد. به عبارت دیگر، گره جدید از تنوع طبقه‌ای کم‌تر از گره مولد خود برخوردار نباشد. در این صورت، این گره به‌عنوان گره برگ درنظرگرفته شده و برچسب عمدهٔ داده‌های متناظر آن به عنوان برچسب طبقه این گره لحاظ می‌شود. در درخت تصمیم دودویی پیاده‌سازی‌شده در این مقاله، انشعاب‌سازی تنها وابسته به مقادیر یک ویژگی بوده که البته بهترین ویژگی ممکن می‌باشد. اگر ویژگی‌ها با x_1، x_2، ... x_p نشان‌داده شوند و p تعداد ویژگی‌ها باشد؛ آن‌گاه داده‌های آموزشی متناظر یک ویژگی نوعی x_k را می‌توان به دو گروه تقسیم کرد، چنان‌که یکی از گروه‌ها دارای مشخصهٔ $x_k < s_k$ و دیگری دارای مشخصهٔ $x_k > s_k$ باشد که s_k آستانهٔ انشعاب است [۴۹].

اگر مقادیر گسستهٔ ویژگی x_k در داده‌های آموزشی متناظر، به صورت افزایشی مرتب شوند، به‌گونه‌ای که $x_k(1) < x_k(2) < \ldots < x_k(M)$ باشد، آن‌گاه میانگین هر دو مقدار متوالی

[DOR: 20.1001.1.23455748.1395.4.1.2.9]

از این مقادیر مرتب‌شده، می‌تواند مطابق رابطهٔ ۱ به عنوان آستانهٔ انشعاب s_k درنظرگرفته شود.

$$T = \frac{x_k(m) + x_k(m+1)}{2} \quad (1)$$

در رابطهٔ ۱، $x_k(m)$ و $x_k(m+1)$ هر دو مقدار متوالی از این مقادیر مرتب‌شده بوده و m =۱, ۲, ... , M-۱ می‌باشد. بنابراین اگر ویژگی x_k دارای M مقدار گسسته باشد، آنگاه باید M-۱ مقدار جهت انتخاب یک آستانهٔ بهینه محاسبه شود. انتخاب آستانهٔ انشعاب بهینه در هر گره تصمیم، با استفاده از معیار انشعاب انجام می‌شود. در این مقاله، از معیار انشعاب جی‌دی‌آی در هر گره مطابق رابطهٔ ۲ استفاده شده است.

$$gdi_{node} = 1 - \sum_i p^2(i) \quad (2)$$

مجموع فوق بر روی طبقه‌های i در گره موردنظر بوده و p(i) نشان‌دهندهٔ کسری از داده‌های متعلق به طبقه i در بین همهٔ داده‌های آموزشی متناظر گره فوق است. یک گره با داده‌های تنها متعلق به یک طبقه، دارای معیار جی‌دی‌آی برابر صفر می‌باشد. واضح است که این گره به عنوان گره برگ درنظرگرفته می‌شود. در حالی که در بقیهٔ حالات، همواره معیار جی‌دی‌آی مثبت و کوچک‌تر از ۱ می‌باشد. در واقع معیار انشعاب، میزان وابستگی یک گره به طبقه‌های مختلف را نشان می‌دهد.

در درخت تصمیم دودویی، یک ویژگی و متعاقباً یک آستانهٔ خاص به گره ریشه تخصیص داده می‌شود؛ برای آن که بهتر از سایر ویژگی‌ها و آستانه‌های محاسبه‌شده برای آن‌ها می‌تواند کل داده‌های آموزشی را به دو گروه تقسیم کرده و معیار جی‌دی‌آی کوچک‌تری تولید کند. به همین ترتیب، به هر گره تصمیم یک ویژگی و یک آستانهٔ انشعاب که معیار جی‌دی‌آی کوچک‌تری را براساس داده‌های متناظر ایجاد نماید، تخصیص می‌یابد.

۲-۱-۲. تشخیص ژانر با استفاده از درخت تصمیم

درخت تصمیم با محاسبهٔ احتمال پسین هر طبقه برای هر قالب از علامت پرس‌وجو، آن را طبقه‌بندی کرده و ژانر آن را تشخیص می‌دهد. واضح است که مجموع این احتمال‌های پسین برای یک قالب برابر ۱ خواهد بود. در واقع برای هر قالب، انشعاب‌های درخت تصمیم با شروع از گره ریشه دنبال شده تا به یک گره برگ برسد. طبقه قالب مطابق رابطه‌های ۳ و ۴، طبقه‌ای است که تابع هزینهٔ طبقه‌بندی را کمینه می‌کند [۴۷]. پس از این‌که طبقه هر قالب از علامت پرس‌وجو توسط درخت تصمیم مشخص شد، با رأی‌گیری حداکثری روی این طبقه‌ها، ژانر نهایی علامت پرس‌وجو مشخص می‌شود.

$$\hat{y} = \sum_{i=1}^{K} P(i \mid Xnew)\, C(k \mid i) \quad (3)$$

$$y = \arg\min_{y=1,\ldots,K} \sum_{i=1}^{K} P(i \mid Xnew) C(y \mid i) \quad (4)$$

که، Xnew یک قالب از علامت پرس‌وجو، $\hat{y}$ هزینهٔ طبقه‌بندی دادهٔ Xnew در طبقه k، y برچسب طبقه تخصیص‌داده‌شده به قالب، K تعداد طبقه‌ها، P(i|Xnew) احتمال پسین طبقه i برای دادهٔ آزمون Xnew و C(y|i) مطابق رابطهٔ ۵، تابع هزینهٔ طبقه‌بندی ناصحیح یک داده در طبقه y است درحالی طبقه صحیح آن i است.

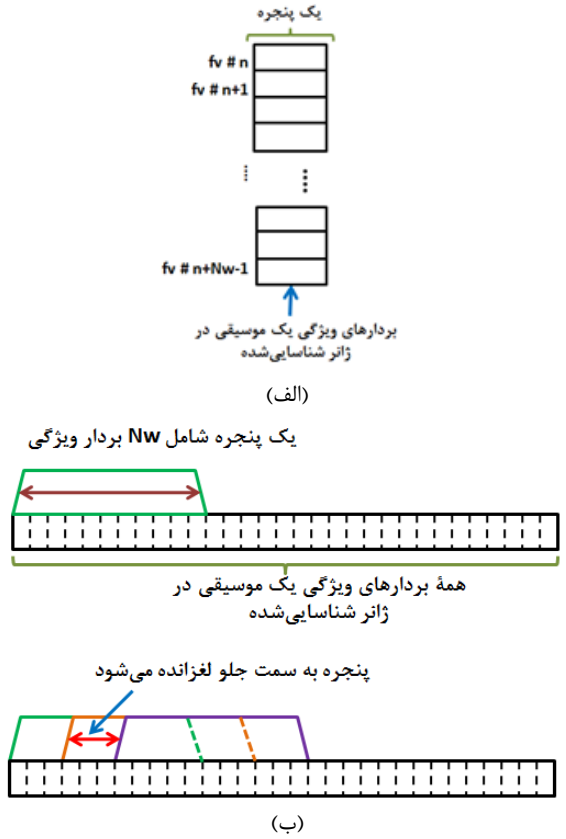
$$cost(y \mid i) = \begin{cases} 1 & \text{if} \quad i \cong y \\ 0 & \text{if} \quad i = y \end{cases} \quad (5)$$

در این مقاله، تعداد کل قالب داده‌ها، N_f، برابر ۱۴۹۹۰۰۰ عدد و تعداد ویژگی‌ها، N_{mfcc}، برابر ۳۰ عدد است. تعداد طبقه‌های دادگان، ۱۰ عدد می‌باشند که به ترتیب از ۱ تا ۱۰ شماره‌گذاری شده‌اند. درخت تصمیم دودویی پیاده‌سازی‌شده دارای ۴۱۸۵۵۹ عدد گره و ۱۹۵۷ لایه است. تعداد قالب‌های علامت پرس‌وجو، N_q، نیز برابر ۲۵۰ عدد می‌باشد.

۲-۲. تطبیق- بازیابی

پس از تشخیص ژانر علامت پرس‌وجو، در طی یک مرحله تطبیق- بازیابی و با استفاده از ترکیب تصمیم مبتنی بر امتیازدهی، آهنگ متناظر علامت پرس‌وجو در طبقه ژانر مربوطه شناسایی و بازیابی می‌شود. این امر با استفاده از یک پنجره لغزان [۱۹] به‌همراه محاسبهٔ یک معیار فاصله و تخصیص امتیاز در هر لغزش انجام می‌شود. در این مقاله، معیار فاصله اقلیدسی [۵] و معیار فاصله واگرایی-کِی‌اِل [۵، ۵۰] مطابق با رابطهٔ ۶ ارزیابی و نتایج آن‌ها با یکدیگر مقایسه شد. ترکیب تصمیم، قانون کمینه را بر روی امتیازهای اخذشده از لغزش پنجره به‌کار می‌گیرد. روند پنجره‌گذاری در شکل ۳ نشان داده شده است.

$$KL\ (p_W(x) \parallel p_Q(x)) = \frac{1}{2}[\log\frac{|\Sigma_Q|}{|\Sigma_W|} + Tr(\Sigma_Q^{-1}\Sigma_W) + (\mu_W - \mu_Q)^T\Sigma_Q^{-1}(\mu_W - \mu_Q) - N_F] \quad (۶)$$



شکل ۳ پنجره‌گذاری و متوالی‌سازی بردارهای ویژگی.

طول پنجرهٔ لغزان، معادل طول علامت پرس‌وجو یعنی ۵ ثانیه بوده و ۲۵۰ عدد قالب را در برمی‌گیرد. N_w تعداد بردارهای ویژگی قرار گرفته در محدودهٔ پنجره و برابر ۲۵۰ است. محل پنجره لغزان از نقطهٔ آغازین هر علامت موسیقی، انتخاب شده و در هر لغزش به اندازهٔ ۰٫۲ ثانیه مطابق شکل ۳-ب به جلو حرکت داده می‌شود که ۱۲۵ لغزش را به ازای هر علامت موسیقی موجب می‌شود. ابتدا، بردارهای ویژگی محتویات قرار گرفته در محدودهٔ پنجره لغزان در یک ماتریس S_w مطابق شکل ۳- الف به صورت متوالی قرار می‌گیرند. سپس همهٔ بردارهای ویژگی علامت پرس‌وجو نیز در یک ماتریس S_Q متوالی شده و فاصلهٔ دو ماتریس S_Q و S_w محاسبه می‌شود. این فاصله به‌عنوان یک امتیاز درنظرگرفته می‌شود. در رابطهٔ ۶، μ_w، μ_Q و Σ_W و Σ_Q به ترتیب میانگین و کوواریانس ماتریس‌های S_Q و S_w هستند و N_F بُعد بردارهای ویژگی است.

۲-۲-۱. ترکیب تصمیم مبتنی بر امتیازدهی

در این مرحله، یک روش ترکیب تصمیم کمینه مبتنی بر امتیازدهی بر روی امتیازهای اخذشده از لغزش پنجره جهت بازیابی آهنگ متناظر با علامت پرس‌وجو مطابق با روابط ۷ تا ۹ پیشنهاد می‌شود. ژانر تشخیص داده‌شده برای علامت پرس‌وجو، شامل ۱۰۰ علامت موسیقی است که لغزش پنجره بر روی این علامت‌های موسیقی منجر به ۱۲۵۰۰ عدد امتیاز حاصل از همهٔ لغزش‌ها می‌شود.

$$score\ (i,j) = distance\ (query, window(s_i, shift_j)) \quad (۷)$$

$$score^*(i) = \min_j (score\ (i,j)) \quad (۸)$$

$$retrieved\ songs\ (r) = \arg\min_{r,i}(score^*\ (i)) \quad (۹)$$

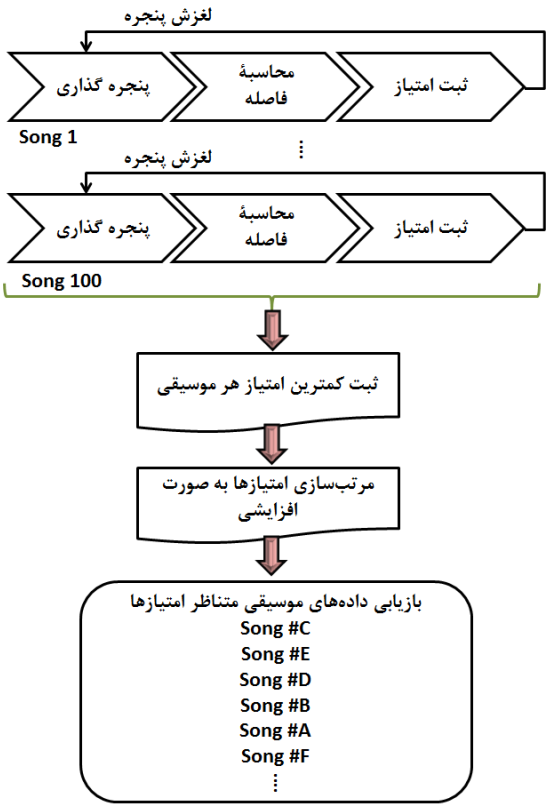
برای i =۱,…,۱۰۰ و j=۱,…,۱۲۵ که s_i یک علامت موسیقی در ژانر تشخیص داده‌شده برای علامت پرس و جو بوده، و $shift_j$ یکی از لغزش‌های پنجرهٔ لغزان است. این روش، تعدادی از شبیه‌ترین علامت‌های موسیقی به علامت پرس‌وجو را بازیابی می‌کند؛ که این علامت‌های موسیقی بر حسب میزان شباهت با علامت پرس‌وجو در خروجی مرتب شده و تعداد آن‌ها با پارامتر «عمق بازیابی»، R، مشخص می‌شوند. در روش فوق، r =۱, …, R می‌باشد. در این راستا، روندنمای بلوک تطبیق- بازیابی در شکل ۴ نشان داده شده است.

۳. نتایج پیاده‌سازی

سامانه بازیابی موسیقی پیشنهادشده در این مقاله از نوع دریافت یک نمونه علامت پرس‌وجو، مبتنی بر تشخیص ژانر و با روی‌کرد بررسی اثر کلی نوفهٔ پس‌زمینه می‌باشد. این روش، با استفاده از دو دسته علامت پرس‌وجوی خالص و نوفه‌ای مورد ارزیابی قرار گرفته‌است. این علامت‌ها به‌صورت تصادفی از دادگان جی‌تی‌زان [۴۰] انتخاب می‌شوند و هر دسته شامل ۱۰۰ علامت پرس‌وجو می‌باشد. طول علامت پرس‌وجو برابر ۵ ثانیه بوده و از لحاظ ژانر، موسیقی متناظر و نقطهٔ شروع، کاملاً تصادفی انتخاب می‌شود. بنابراین پس از قالب‌بندی علامت

[DOR: 20.1001.1.23455748.1395.4.1.2.9]

پرس‌وجو، هیچ الزامی مبنی بر انطباق کامل آن بـر فضـای آموزشی وجود ندارد.



شکل ۴ روندنمای بلوک تطبیق- بازیابی.

۳-۱. نتایج بلوک تشخیص ژانر

در ایـن بخـش، نتـایج درخـت تصـمیم دودویـی پیاده‌سازی‌شده در بلوک تشخیص ژانر، ارزیـابی مـی‌شـود. این ارزیابی، از طریق ماتریس سـرگردانی و محاسـبهٔ بـازهٔ خطا [۵۱] انجام می‌گیرد. جدول‌های ۱ و ۲ مـاتریس‌هـای سرگردانی را به ترتیب برای دو دسته علامت پـرس‌وجـوی خالص و نوفه‌ای نشان می‌دهند. مـاتریس‌هـای سـرگردانی تقریباً قطری هستند که نشان‌دهندهٔ قدرت درخت تصمیم دودویی در تشخیص ژانر علامت پرس‌وجو می‌باشد.

بازهٔ خطا برای دو حالت خالص و نوفـه‌ای بـه ترتیـب ۳٪ و ۱۴٪ است. نتایج جـدول‌هـای ۱ و ۲ نشـان مـی‌دهـد کـه اگرچه تمرکز خطا روی ژانر خاصی وجود ندارد؛ ولی میزان خطا در تشخیص ژانر در حالت وجود نوفه بـرای ژانرهـای بولوز، دیسکو و راک نسـبت بـه حالـت عـدم وجـود نوفـه افزایش یافته است.

جدول ۱ ماتریس سرگردانی برای دسته علامت‌های پرس‌وجوی خالص.

ژانر شناسایی‌شده												
		بولوز	کلاسیک	کانتری	دیسکو	هیپ‌هاپ	جاز	متال	پاپ	رگا	راک
ژانر معلوم	بولوز	۱۰	۰	۰	۰	۰	۱	۰	۰	۰	۰
	کلاسیک	۰	۱۲	۰	۰	۰	۰	۰	۰	۰	۰
	کانتری	۰	۰	۱۰	۰	۰	۰	۰	۰	۰	۰
	دیسکو	۰	۰	۰	۱۳	۰	۰	۰	۰	۰	۰
	هیپ‌هاپ	۰	۰	۰	۰	۹	۰	۰	۰	۰	۰
	جاز	۰	۱	۰	۰	۰	۷	۰	۰	۰	۰
	متال	۰	۰	۱	۰	۰	۰	۱۰	۰	۰	۰
	پاپ	۰	۰	۰	۰	۰	۰	۰	۱۱	۰	۰
	رگا	۰	۰	۰	۰	۰	۰	۰	۰	۸	۰
	راک	۰	۰	۰	۰	۰	۰	۰	۰	۰	۷

جدول ۲ ماتریس سرگردانی برای دسته علامت‌های پرس‌وجوی نوفه‌ای.

ژانر شناسایی‌شده												
		بولوز	کلاسیک	کانتری	دیسکو	هیپ‌هاپ	جاز	متال	پاپ	رگا	راک
ژانر معلوم	بولوز	۷	۰	۰	۰	۰	۴	۰	۰	۰	۰
	کلاسیک	۰	۱۱	۰	۰	۰	۱	۰	۰	۰	۰
	کانتری	۰	۰	۹	۰	۰	۰	۰	۰	۱	۰
	دیسکو	۰	۰	۱	۱۰	۰	۰	۰	۰	۲	۰
	هیپ‌هاپ	۰	۰	۰	۰	۸	۰	۰	۱	۰	۰
	جاز	۰	۰	۰	۰	۰	۸	۰	۰	۰	۰
	متال	۰	۰	۱	۰	۰	۰	۱۰	۰	۰	۰
	پاپ	۰	۰	۰	۰	۰	۰	۰	۱۱	۰	۰
	رگا	۰	۰	۰	۰	۰	۰	۰	۰	۸	۰
	راک	۰	۰	۰	۰	۱	۲	۰	۰	۰	۴

در این‌جـا لازم اسـت تاکیـد شـود کـه در یـک الگـوریتم تشخیص ژانـر صـرف، داده‌هـای آمـوزش و آزمـایش بایـد به‌طور کامل از یکدیگر مجزا باشند، در حالی‌که روش ارائـه شده در این مقالـه چنـین نیسـت؛ و داده‌هـای آمـوزش و آزمایش کـاملاً از یکـدیگر جـدا نبـوده و بـه‌نـوعی تصـویر علامت پرس‌وجو (نـه لزومـاً عـین علامـت پـرس‌وجـو) در فضای آموزشی وجود دارد.

[DOR: 20.1001.1.23455748.1395.4.1.2.9]

همان‌گونه که قبلاً نیز اشاره شد، بلوک تشخیص ژانر در سامانه پیشنهادی صرفاً یک عامل کمکی برای کاهش فضای جستجو و افزایش سرعت سامانه است؛ و نحوهٔ خاص پیاده‌سازی این بلوک یعنی عدم جداسازی کامل فضای آموزش و آزمایش از یکدیگر نیز در راستای نیل به این هدف می‌باشد.

به عبارت دیگر، هدف دنبال‌شده در این مقاله، پیشنهاد یک الگوریتم با کارایی بالا در مقولهٔ تشخیص ژانر نبوده‌است؛ که در این صورت، به‌طورمنطقی باید فضای آموزش و آزمایش کاملاً از یکدیگر جدا بوده و از روش‌هایی مشخصی مانند کِی- فولد[1] و اعتبارسنجی مقطعی (کراس ولیدیشن)[2] برای بیان نتایج استفاده شود. این در حالی است که در این مقاله، صرف ارزیابی عملکرد درخت تصمیم دودویی در تشخیص درست ژانر علامت پرس‌وجو برای سامانه پیشنهادی مد نظر بوده است. هم‌چنین در این کار فرض بر این است که بلوک تشخیص ژانر با آهنگ جدیدی که در دادگان وجود ندارد مواجه نمی‌شود و دادگان نیز تغییر نمی‌کند که در صورت افزوده شدن آهنگ جدیدی به دادگان، بلوک تشخیص ژانر باید روزآمد شود.

بنابراین، جهت بیان عدم کارایی کافی الگوریتم کارت در شرایط جدا بودن فضای آموزش و آزمایش، این الگوریتم در دو حالت ۳- فولد کراس ولیدیشن و ۱۰- فولد کراس ولیدیشن، با استفاده از دادگان جی‌تی‌زان و جهت نشان دادن تفاوت بین این آزمون‌ها و روش استفاده شده در این مقاله، مورد ارزیابی قرار گرفت؛ که نتایج نشان داد نرخ طبقه‌بندی ژانر به ترتیب برای دو حالت فوق ۶۸٫۲۰٪ و ۷۲٫۳۵٪ است.

نتایج فوق را می‌توان با نرخ ۹۵٫۷٪ برای طبقه‌بندی ژانر دادگان جی‌تی‌زان، که به‌وسیلهٔ نمایش تُنُک از ویژگی‌های بسآمدی و در حالت ۵- فولد کراس ولیدیشن [۵۲] حاصل شده، مقایسه کرد. هم‌چنین دادگان فوق در منبع [۳۲] نیز مورد استفاده قرار گرفته‌اند و نرخ ۴±۶۱٪ برای طبقه‌بندی ژانر این دادگان با استفاده از الگوی گوسی با ۵ مخلوط و در حالت ۱۰- فولد کراس ولیدیشن به‌دست آمده‌اند.

۳-۲. نتایج بلوک تطبیق- بازیابی

نتایج سامانه بازیابی موسیقی پیشنهادی با پارامترهای صحت، زمان بازیابی و نمودار دقت- فراخوان سنجیده می‌شود. رابطهٔ ۱۰ پارامتر صحت را تعریف می‌کند [۳۲، ۵۳]:

$$A = \frac{c}{q} \qquad (۱۰)$$

که در آن، A میزان صحت، c تعداد داده‌های صحیح بازیابی‌شده و q تعداد علامت‌های پرس‌وجو می‌باشند. تعداد داده‌های موسیقی بازیابی‌شده که در خروجی ارائه می‌شوند با پارامتر «عمق بازیابی» مشخص شده‌اند. واضح است که تنها یکی از این داده‌های بازیابی‌شده در خروجی، ممکن است موسیقی مطلوب کاربر باشد. در صورتی که موسیقی مطلوب کاربر در بین داده‌های ارائه‌شده در خروجی وجود داشته باشد، این خروجی به عنوان یک نتیجه صحیح لحاظ می‌شود.

جدول‌های ۳ و ۴ میزان صحت نتایج بازیابی را برای دو دسته علامت پرس‌وجوی خالص و نوفه‌ای در عمق بازیابی ۱ تا ۱۰ به ترتیب برای فاصلهٔ اقلیدسی و کِی‌اِل نشان می‌دهند. نتیجهٔ صحیح بیان‌گر این است که هم ژانر علامت پرس‌وجو به درستی تشخیص داده شده، و هم آهنگ متناظر به درستی مورد بازیابی قرار گرفته است. در عمق بازیابی ۱ تنها یک دادهٔ موسیقی، بازیابی شده و در خروجی ارائه می‌شود؛ که ممکن است موسیقی مطلوب کاربر باشد یا نباشد. در عمق بازیابی بیش از ۱، تعدادی دادهٔ موسیقی در خروجی ارائه می‌شود که این داده‌ها برحسب میزان شباهت به علامت پرس‌وجو مرتب شده‌اند، و آهنگ مطلوب کاربر، ممکن است در میان آن‌ها باشد یا نباشد.

از جدول‌های ۳ و ۴ واضح است که تقریباً در عمق بازیابی ۳، یعنی ارائهٔ سه عدد دادهٔ موسیقی بازیابی‌شده در خروجی، می‌توان مطلوب‌ترین حالت ممکن از نتایج را به‌دست آورد. این امر در ادامه، توسط نمودار دقت- فراخوان نیز تایید می‌شود. البته نتایج در جدول ۴ نشان می‌دهند که عملکرد فاصلهٔ کِی‌اِل در حالت خالص بهتر از فاصلهٔ اقلیدسی می‌باشد. در حالی که فاصلهٔ اقلیدسی در

[1] K-fold
[2] Cross validation

[DOR: 20.1001.1.23455748.1395.4.1.2.9]

حالت نوفه‌ای مقاوم‌تر بوده و عملکرد بهتری نشان داده است.

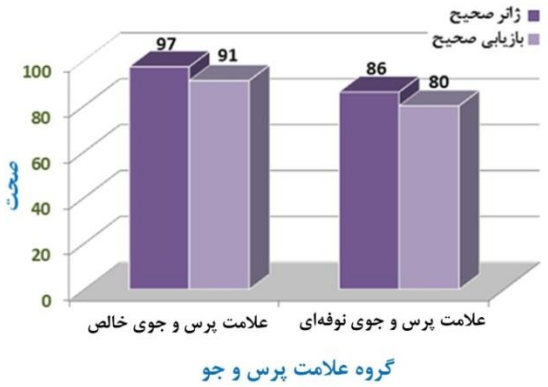
جدول ۳ میزان صحت در عمق‌های بازیابی مختلف با استفاده از فاصله اقلیدسی.

عمق بازیابی	حالت خالص		حالت نوفه‌ای (اِس‌اِن‌آر= ۳۰ دسی‌بل)	
	صحت بازیابی (٪)	تشخیص صحیح ژانر (٪)	صحت بازیابی (٪)	تشخیص صحیح ژانر (٪)
۱	۹۱	۹۷	۸۰	۸۶
۲	۹۶		۸۴	
۳	۹۷		۸۶	
۴	۹۷		۸۶	
۵	۹۷		۸۶	
۶	۹۷		۸۶	
۷	۹۷		۸۶	
۸	۹۷		۸۶	
۹	۹۷		۸۶	
۱۰	۹۷		۸۶	

جدول ۴ میزان صحت در عمق‌های بازیابی مختلف با استفاده از فاصله واگرایی-کِی‌اِل.

عمق بازیابی	حالت خالص		حالت نوفه‌ای (اِس‌اِن‌آر= ۳۰ دسی‌بل)	
	صحت بازیابی (٪)	تشخیص صحیح ژانر (٪)	صحت بازیابی (٪)	تشخیص صحیح ژانر (٪)
۱	۹۵	۹۷	۵۸	۸۶
۲	۹۷		۶۲	
۳	۹۷		۶۳	
۴	۹۷		۶۷	
۵	۹۷		۷۰	
۶	۹۷		۷۱	
۷	۹۷		۷۵	
۸	۹۷		۸۰	
۹	۹۷		۸۲	
۱۰	۹۷		۸۲	

شکل ۵ نتایج سامانه بازیابی موسیقی پیشنهادی را برای دو دسته علامت پرس‌وجوی خالص و نوفه‌ای در عمق بازیابی ۱ برای فاصلهٔ اقلیدسی نشان می‌دهد. پارامتر «ژانر صحیح» نشان‌دهندهٔ تعداد علامت‌های پرس‌وجویی است که ژانر آن‌ها به‌درستی تشخیص داده شده‌اند. پارامتر «بازیابی صحیح» نشان‌دهنده تعداد علامت‌های پرس‌وجویی است که به درستی بازیابی شده‌اند. به عبارت دیگر، تنها دادهٔ موسیقی ارائه‌شده در خروجی همان موسیقی مطلوب کاربر بوده است.



شکل ۵ نتایج سامانه پیشنهادی در عمق بازیابی ۱ برای فاصلهٔ اقلیدسی.

نتایج ارائه‌شده در شکل ۵، دستیابی به میزان صحت ۹۱٪ را در عمق بازیابی ۱ برای دستهٔ علامت پرس‌وجوی خالص نشان می‌دهد. در مقابل، نتایج، دستیابی به میزان صحت ۸۰٪ در عمق بازیابی ۱ برای دستهٔ علامت پرس‌وجوی نوفه‌ای را تایید می‌کند. البته تاکید می‌شود که نتایج در شکل ۵ برای عمق بازیابی ۱ است. یعنی تنها یک دادهٔ موسیقی در خروجی بازیابی می‌شود؛ که می‌تواند موسیقی مطلوب کاربر باشد یا نباشد. واضح است، که برای عمق بازیابی بزرگ‌تر از ۱ و ارائهٔ یک مجموعه دادهٔ موسیقی در خروجی، شانس وجود موسیقی مطلوب کاربر در بین آن‌ها بیش‌تر خواهد بود که مسلماً منجر به افزایش پارامتر صحت خواهد شد.

زمان بازیابی متناظر دو دسته علامت پرس‌وجوی فوق برابر ۵۲۵ میلی‌ثانیه برای فاصلهٔ اقلیدسی، و ۳۸۰ میلی‌ثانیه برای فاصلهٔ کِی‌اِل می‌باشد. البته این زمان در واقع مدت زمانی است که بار محاسباتی تا حصول نتیجه به سی‌پی‌یو[1] اعمال می‌شود. مسلماً در یک سامانه کاربردی، زمان لازم جهت انتقال نتیجه از حافظه و نمایش آن به زمان فوق افزوده می‌شود.

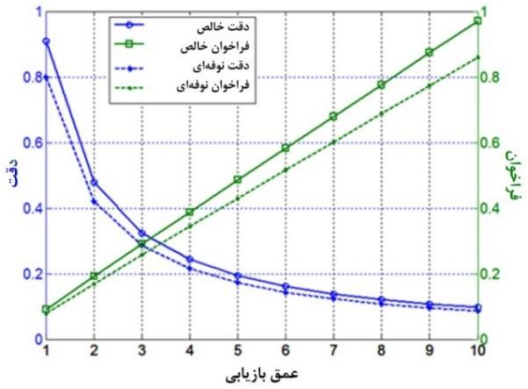
یک نمودار شناخته‌شده برای ارزیابی عملکرد یک سامانه بازیابی اطلاعات، نمودار دقت- فراخوان[2] است. روابط ۱۱ و ۱۲ به ترتیب پارامتر فراخوان و دقت را بیان می‌کنند.

[1] CPU
[2] Recall-precision

$$\text{فراخوان} = \frac{\text{تعداد تموته‌های متناسب که بازیابی شده‌اند}}{\text{تعداد کل تموته‌های متناسب}} \quad (۱۱)$$

$$\text{دقت} = \frac{\text{تعداد تموته‌های متناسب که بازیابی شده‌اند}}{\text{تعداد کل تموته‌های بازیابی شده}} \quad (۱۲)$$

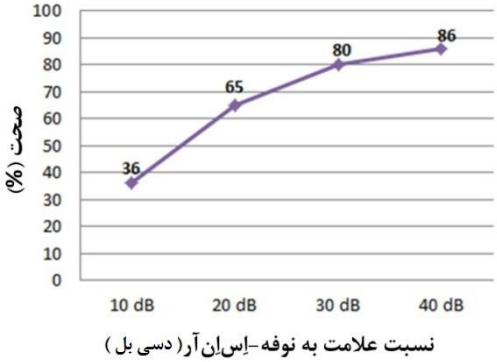
شکل ۶ نمودار دقت- فراخوان را برای سامانه بازیابی موسیقی پیشنهاد شده در این مقاله بر حسب عمق‌های بازیابی مختلف و برای دو دسته علامت پرس‌وجوی خالص و نوفه‌ای با استفاده از فاصلهٔ اقلیدسی نشان می‌دهد.



شکل ۶ نمودار دقت- فراخوان علامت‌های پرس‌وجوی خالص و نوفه‌ای.

واضح است که حتی برای عمق بازیابی بزرگ‌تر از یک و ارائه بیش از یک دادهٔ موسیقی در خروجی، تنها یکی از این داده‌ها ممکن است موسیقی مطلوب کاربر باشد. بنابراین، طبق تعریف پارامتر دقت، افزایش عمق بازیابی، کاهش دقت را به دنبال دارد. در واقع، به‌نوعی شانس بازیابی موسیقی مطلوب کاربر بین سایر داده‌های بازیابی‌شده تقسیم می‌شود. از سوی دیگر، افزایش عمق بازیابی منجر به بازیابی سایر موسیقی‌های مشابه موسیقی مطلوب کاربر نیز می‌شود. لذا با توجه به تعریف پارامتر فراخوان، افزایش این پارامتر را به‌همراه خواهد داشت. شکل ۶ تایید می‌کند که با توجه به تعاریف پارامترهای دقت و فراخوان، عمق بازیابی ۳، انتخابی مناسب برای سامانه بازیابی موسیقی پیشنهادشده در این مقاله می‌باشد. شکل ۷ نتایج ارزیابی سامانه بازیابی موسیقی با نسبت‌های مختلف علامت به نوفه را در عمق بازیابی ۱ برای فاصلهٔ اقلیدسی نشان می‌دهد. همان‌گونه که از شکل ۷ نیز واضح است؛ با افزایش میزان نوفه سفید گوسی (و در نتیجه کاهش نسبت علامت به نوفه)، علامت پرس‌وجوی ورودی به سامانه دارای کیفیت مطلوبی نبوده و بنابراین احتمال یافتن و بازیابی آهنگ متناظر آن کاهش می‌یابد.

معمولاً در پردازش علامت‌های موسیقی، تعداد ضرایب کپسترال مِل استفاده شده بیش از ۹ عدد (بسته به نوع پژوهش) بوده و در برخی کاربردهای مرتبط با زمینهٔ موسیقی، تا ۲۰ یا ۳۰ عدد نیز انتخاب می‌شود. بر این اساس، عملکرد سامانه بازیابی موسیقی پیشنهاد شده در این مقاله به‌صورت تجربی در حالت‌های استفاده از ۱۰، ۱۵، ۲۰ و ۳۰ ضریب کپسترال مِل به عنوان ویژگی آزمایش شد. نتایج آزمایش نشان داد عملکرد سامانه پیشنهادی با افزایش تعداد ضرایب کپسترال مِل، بهبود می‌یابد؛ ولی این بهبود بین دو حالت استفاده از ۲۰ ضریب و ۳۰ ضریب کم‌تر است. در نهایت با توجه به نتایج تجربی بدست آمده مطابق با جدول ۵، استفاده از ۳۰ ضریب کپسترال مِل به عنوان ویژگی در سامانه بازیابی موسیقی پیشنهادی انتخاب شد.



شکل ۷ نمودار صحت بر حسب نسبت علامت به نوفه در عمق بازیابی ۱ برای فاصلهٔ اقلیدسی.

جدول ۵ نتایج تجربی برای انتخاب تعداد ضرایب کپسترال مِل.

با فرض درست بودن ژانر و فاصله اقلیدسی در عمق بازیابی ۱			
۳۰اِم‌اِف‌سی‌سی	۲۰اِم‌اِف‌سی‌سی	۱۵اِم‌اِف‌سی‌سی	۱۰اِم‌اِف‌سی‌سی
٪۹۴	٪۹۲	٪۸۷	٪ ۸۳

۳-۳. مقایسه با روش جستجوی پایه

هدف از پیاده‌سازی روش جستجوی پایه، مقایسه نتایج بازیابی در دو حالت وجود و عدم وجود بلوک تشخیص ژانر به عنوان بلوک پیشنهادی جهت کاهش فضای جستجو و

کاهش زمان بازیابی است. روش جستجوی پایه تنها با استفاده از لغزش پنجرهٔ لغزان بر روی کل فضای آموزشی و محاسبهٔ فاصلهٔ اقلیدسی در هر موقعیت جهت بازیابی داده‌های موسیقی مشابه علامت پرس‌وجو انجام می‌شود. جدول ۶ شامل نتایج روش جستجوی پایه در عمق بازیابی ۱ و با استفاده از ۳۰ ضریب کپسترال مِل می‌باشد.

جدول ۶ نتایج روش جستجوی پایه برای دو دسته علامت پرس‌وجوی خالص و نوفه‌ای.

		عمق بازیابی ۱	
		حالت خالص	حالت نوفه‌ای (اِس‌اِن‌آر= ۳۰ دسی‌بل)
ضرایب کپسترال مِل	تشخیص صحیح ژانر (٪)	۹۸	۹۰
	صحت بازیابی (٪)	۹۲	۸۸

نتایج جدول ۶ نیز نشان‌دهندهٔ میزان صحت به ترتیب ۹۲٪ و ۸۸٪ برای دو دسته علامت پرس‌وجوی خالص و نوفه‌ای است. در حالی که زمان بازیابی در روش جستجوی پایه برابر ۵٫۲ ثانیه می‌باشد. واضح است که چنین زمان بازیابی برای پیاده‌سازی یک سامانه بازیابی موسیقی مطلوب، کاربردی و نزدیک به بلادرنگ مناسب نیست.

اگر چه سامانه پیشنهادی، موفقیت کم‌تری در حالت نوفه‌ای نسبت به روش جستجوی پایه حاصل کرده است، اما کسب موفقیت ۸۶٪ در بازیابی حالت نوفه‌ای توسط سامانه بازیابی موسیقی پیشنهادی و کاهش زمان بازیابی از ۵٫۲ ثانیه به ۵۲۵ میلی‌ثانیه هم‌چنان مطلوب است. علاوه بر این‌که نشان می‌دهد تشخیص ژانر می‌تواند در پیاده‌سازی یک سامانه بازیابی موسیقی سریع و دقیق مؤثر باشد.

۳-۴. مقایسه با افزودن ویژگی‌های زمان- بسآمد

هدف از این بخش، ارزیابی تاثیر افزودن تعدادی ویژگی شناخته‌شدهٔ زمانی- بسآمدی به ضرایب کپسترال مِل در بلوک تطبیق- بازیابی است. این ویژگی‌ها عبارتند از: نرخ عبور از صفر، بیشینه دامنه زمانی هر قاب، کمینه دامنهٔ زمانی هر قاب، انرژی کل در طیف زمانی، بیشینه اول و دوم در انرژی محلی طیف زمانی، مرکز ثقل طیف بسآمدی، شار طیف بسآمدی و نقطه دوران انرژی طیف. مجموع ویژگی‌های فوق در کنار ضرایب کپسترال مِل، بردار ویژگی با طول ۳۹ را به‌دنبال دارد. تاکید می‌شود که در این حالت هم‌چنان تشخیص ژانر با استفاده از ۳۰ ضریب کپسترال مِل انجام شده و تنها بلوک تطبیق- بازیابی با افزودن ویژگی‌های فوق ارزیابی می‌شود. جدول ۷ نتایج این ارزیابی را با فاصلهٔ اقلیدسی نشان می‌دهد. زمان بازیابی در این حالت، حدود ۱ ثانیه است.

جدول ۷ میزان صحت در عمق‌های بازیابی مختلف با افزایش ویژگی‌ها.

عمق بازیابی	حالت خالص		حالت نوفه‌ای (اِس‌اِن‌آر= ۳۰ دسی‌بل)	
	تشخیص صحیح ژانر (٪)	صحت بازیابی (٪)	تشخیص صحیح ژانر (٪)	صحت بازیابی (٪)
۱	۹۷	۹۰	۸۶	۷۹
۲		۹۵		۸۴
۳		۹۵		۸۵
۴		۹۷		۸۶
۵		۹۷		۸۶
۶		۹۷		۸۶
۷		۹۷		۸۶
۸		۹۷		۸۶
۹		۹۷		۸۶
۱۰		۹۷		۸۶

مقایسه نتایج در جدول ۷ با نتایج در جدول ۳ نشان می‌دهد که افزودن ویژگی‌های زمانی- بسآمدی نه تنها موفقیت بیش‌تری در بازیابی را به‌دنبال نداشته، بلکه باعث افت نتایج در عمق‌های بازیابی ۱ تا ۳ نیز شده‌اند. این در حالی است که افزودن این ویژگی‌ها، افزایش هزینهٔ محاسباتی و در نتیجه افزایش زمان بازیابی از ۵۲۵ میلی‌ثانیه به حدود ۱ ثانیه را نیز به‌همراه دارد. ارزیابی فوق مؤید این نکته است که لزوماً تعدد ویژگی‌ها یا تنوع ویژگی‌ها در یک روش بازیابی، شرط لازم برای دستیابی به نتایج بازیابی بهتر همراه با زمان بازیابی مناسب نیست. بلکه، کلید موفقیت یک سامانه بازیابی موسیقی، انتخاب بهینه‌ترین ویژگی‌هایی است که هم نتایج مطلوب بازیابی را در پی داشته باشد، و هم موجب افزونگی محاسباتی نشده و حصول نتایج در زمان بازیابی مناسب را محقق سازد.

[DOR: 20.1001.1.23455748.1395.4.1.2.9]

۳-۵. مقایسه با چند سامانه دیگر

بدیهی است، سامانه‌های کامل بازیابی موسیقی، تنها به بخشی که در این مقاله مطرح شد، محدود نمی‌شوند. آن‌ها از فراداده‌های زیادی استفاده می‌کنند که بسیار مؤثر هستند. به‌عنوان مثال می‌توان به فهرست آهنگ‌های پرطرفدار در بازه‌های زمانی مختلف در هر کشور یا منطقه جغرافیایی، اطلاعات با ارزش حاصل از داده‌کاوی حجم عظیم پرس‌وجوهای دریافتی، اطلاعات شخصی کاربر و سابقه او در استفاده از سامانه اشاره کرد. هم‌چنین، بازخورد ربط[1] که ممکن است از کاربر بگیرند، در نتیجهٔ بازیابی بسیار مؤثر خواهد بود.

در کارهای پژوهشی غیرتجاری نیز، این موضوع که روش ارائه‌شده تا چه حد به یک روش کامل بازیابی نزدیک است، در کارایی آن بسیار مؤثر خواهد بود. حجم بانک داده و تنوع موسیقیایی آن هم در کارهای مختلف، گوناگون است. بنابراین آنچه در ادامه می‌آید با در نظر داشتن این مسائل قابل بررسی می‌باشد.

در اولین گام، عملکرد سامانه پیشنهادی به‌طورنسبی با چند نمونه نرم‌افزار کاربردی مقایسه می‌شود. البته، واضح است که به دلیل تجاری بودن این‌گونه نرم‌افزارها، اطلاعات دقیقی راجع به پیش فرض‌ها و روش به‌کاررفته در این نرم‌افزارها و هم‌چنین دادگان استفاده شده در آن‌ها در دسترس نیست و تنها می‌توان یک مقایسهٔ نسبی را با آن‌ها انجام داد.

در منبع [۵۴]، یک ارزیابی بین دو سامانه شازام [۳۱] و سوندهوند [۳۵] که هر دو از جمله نرم‌افزارهای تجاری گوشی همراه برای بازیابی موسیقی بر مبنای نمونه هستند، انجام گرفته است. نتایج این ارزیابی در جدول ۸ قرار دارند. براساس این نتایج، سامانه شازام سریع‌تر از سوندهوند می‌باشد؛ در حالی‌که میزان صحت سوندهوند بیش از شازام است. به‌طورکلی به نظر می‌رسد، که میزان صحت ۹۵٪ در سامانه‌های بازیابی موسیقی مبتنی بر نمونه برای دستیابی به مطلوبیت تجاری قابل رقابت کفایت می‌کند. اگر چه نرخ خطای ۵٪ هم‌چنان برای تایید و تصدیق چنین سامانه‌هایی بالا می‌باشد [۵۵]. در جدول ۸ مشخصات عملکردی سامانهٔ پیشنهادی جهت مقایسهٔ نسبی آورده شده است.

جدول ۸ مقایسهٔ نسبی بین سامانه پیشنهادی و نرم‌افزارهای تجاری شازام و سوندهوند.

سامانه	عمق بازیابی	حالت خالص		
		صحت بازیابی (٪)	زمان بازیابی	شرایط آزمایش
شازام [۳۱]	۱	۸۵	۱۲ ثانیه	بر روی گوشی‌همراه با سامانهٔ عامل اندروید
سوندهوند [۳۵]	۱	۹۵	۲۱ ثانیه	بر روی گوشی‌همراه با سامانهٔ عامل اندروید
سامانه پیشنهادی با فاصلهٔ کِی‌اِل	۱	۹۵	۳۸۰ میلی ثانیه	زمان احتسابی برای سی‌پی‌یو روی GHz ۳٫۸ آی۸‌پی‌سی و با کد متلب
سامانه پیشنهادی با فاصلهٔ اقلیدسی	۱	۹۱	۵۲۵ میلی ثانیه	زمان احتسابی برای سی‌پی‌یو روی GHz ۳٫۸ آی۸‌پی‌سی و با کد متلب

هم‌چنین مقاومت نرم‌افزار شازام در مقابل نوفهٔ پس‌زمینه در منبع [۳۷] آزمایش شده است. برای انجام این آزمایش، مقادیر مختلف نوفه سفید گوسی به علامت‌های پرس‌وجو افزوده می‌شوند. نتایج این آزمایش، نرخ بازیابی صحیح ۵۰٪ برای علامت‌های پرس‌وجوی ۱۰ ثانیه‌ای در حالت نسبت علامت به نوفهٔ صفر دسی‌بل را نشان می‌دهند. در حالتی‌که نسبت علامت به نوفه به ۶ دسی‌بل افزایش یابد، نرخ بازیابی صحیح به حدود ۹۰٪ می‌رسد.

مقایسهٔ مقادیر فوق با نتایج ارائه شده در شکل ۷ نشان می‌دهند که، عملکرد سامانه بازیابی موسیقی پیشنهادشده در این مقاله در مقابل مقادیر معمول نوفه (نسبت علامت به نوفه بین ۳۰ تا ۴۰ دسی‌بل)، تقریباً مقاوم است.

جهت مقایسهٔ عملکرد سامانه پیشنهادی در این مقاله با چند نمونه سامانه دیگر که در مقدمه معرفی شدند، دو

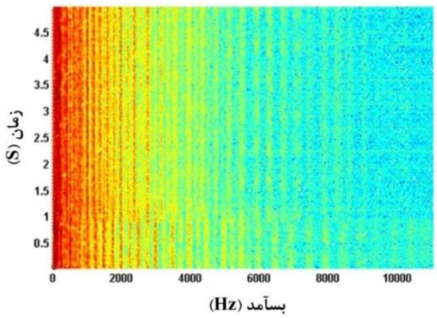
[1] Relevance feedback

[DOR: 20.1001.1.23455748.1395.4.1.2.9]

روش زیر پیاده‌سازی می‌شوند. در روش اول، مشابه توضیحاتی که برای سامانه شازام در مقدمه ذکر شد، از روش اثر انگشت[1] و شناسایی بیشینه‌های طیف بسآمد- زمان برای ایجاد یک مجموعه ویژگی تُنُک در یک بازهٔ بسآمدی خاص استفاده می‌شود. در روش دوم، علامت به‌وسیله الگوی ترکیب گوسی پارامتری شده و این پارامترها برای جستجو در دادگان با استفاده از یک معیار فاصله به‌کارگرفته می‌شوند [۵].

۳-۵-۱. روش اول

در این روش، تقریباً مشابه سامانه شازام، یک پنجرهٔ لغزان بر روی کل فضای آموزشی حرکت داده‌شده و در هر موقعیت که به آن اثر انگشت نیز گفته می‌شود، تعدادی از بیشینه‌های طیف بسآمد- زمان در یک بازهٔ بسآمدی خاص شناسایی شده و به‌عنوان ویژگی‌های متناظر هر اثر انگشت ذخیره می‌شوند. سپس از این ویژگی‌ها برای محاسبهٔ فاصله تا علامت پرس‌وجو، و در نهایت بازیابی دادهٔ موسیقی مشابه با آن استفاده می‌شود. شکل ۸، طیف بسآمد- زمان را متعلق به یکی از اثر انگشت‌ها نشان می‌دهد.



شکل ۸ طیف بسآمدی- زمانی متعلق به یک قطعهٔ ۵ ثانیه‌ای از یک علامت موسیقی نوعی در دادگان جی‌تی‌زان [۴۰].

بررسی نشان داد که برای اثر انگشت‌ها، تمرکز انرژی عمدتاً تا بسآمد ۵ کیلوهرتز می‌باشد. بنابراین، در این پیاده‌سازی، ۵۰ بسآمد مشخص با فاصلهٔ ۱۰۰ هرتز از یکدیگر در بازهٔ بسآمدی فوق انتخاب شده و بیشینهٔ متناظر هر بسآمد شناسایی شد. لذا، به ازای هر اثر انگشت، یک بردار ویژگی ۵۰ مؤلفه‌ای به‌دست می‌آید. پارامترهای طیف بسآمد- زمان به‌گونه‌ای تنظیم شده‌اند که اثر انگشت مربوطه، به بخش‌هایی با طول زمانی ۴۰ میلی‌ثانیه و با هم‌پوشانی ۵۰٪ (مشابه فرایند قالب‌بندی) تقسیم شود. جدول ۹ نتایج این پیاده‌سازی را با استفاده از فاصلهٔ اقلیدسی نشان می‌دهد. روش انتخاب اثر انگشت و مشخصهٔ آن در سامانه شازام مبتنی بر الگوریتم هاشینگ[2] [۵۶] می‌باشد. البته همان‌گونه که قبلاً ذکر شد، به دلیل تجاری بودن سامانه شازام، اطلاع دقیقی از مشخصات روش آن در مورد انتخاب و اولویت‌بندی اثر انگشت‌ها، بازهٔ بسآمدی استفاده‌شده، تعداد بسآمدهای انتخابی، نقاط بسآمدی بهینه و نحوهٔ محاسبهٔ فاصله و انطباق در دسترس نیست.

۳-۵-۲. روش دوم

در این روش مشابه روش جستجوی پایه، یک پنجرهٔ لغزان بر روی کل فضای آموزشی حرکت داده‌شده و در هر موقعیت، قالب‌بندی و استخراج ویژگی‌های کپسترال مِل انجام می‌شود. سپس مشابه کار انجام‌شده در منبع [۵]، توزیع بردارهای ویژگی در هر موقعیت با یک الگوی گوسی شامل ۱۲ ترکیب تخمین زده شده و پارامترهای الگو ذخیره می‌شوند. این پارامترها شامل وزن، میانگین و واریانس ترکیب‌های موجود در الگوی گوسی برای هر موقعیت می‌باشند. سپس از این پارامترها برای محاسبهٔ فاصله تا علامت پرس‌وجو و در نهایت بازیابی دادهٔ موسیقی مشابه با آن استفاده می‌شود. در منبع [۵]، یک شکل بسته[3] برای محاسبهٔ فاصلهٔ اقلیدسی بین دو الگوی ترکیب گوسی پیشنهاد شده است، و از آن‌جا که هیچ شکل بسته‌ای برای محاسبهٔ فاصلهٔ کِی‌اِل بین دو الگوی گوسی با بیش از یک ترکیب وجود ندارد، در منبع [۵] تعدادی از شکل‌های تقریبی که در سال‌های اخیر برای این منظور پیشنهاد شده، بررسی و آزمایش شده‌اند، که همگی بار محاسباتی زیادی دارند. بنابراین، در این‌جا از روش پیشنهادشده در منبع [۵۷] برای اندازه‌گیری فاصلهٔ بین دو الگوی ترکیب گوسی استفاده می‌شود. نتایج این پیاده‌سازی با استفاده از فاصلهٔ فوق در جدول ۹ بیان شده است.

[1] Fingerprint

[2] Hashing
[3] Closed form

البته روش بکار رفته در منبع [۵] بر روی یک دادگان از انواع صدا (گفتار، موسیقی، آواز و صداهای محیطی شامل صدای داخل اتومبیل، صدای رستوران و صدای جاده) و به منظور پیاده‌سازی یک سامانه طبقه‌بندی صدا بر مبنای نمونه بکار رفته است. در این سامانه، یک نمونه علامت پرس‌وجو توسط سامانه دریافت شده و گروه صوتی آن مشخص می‌شود. بنابراین به دلیل تفاوت نوع دادگان و هم‌چنین تفاوت نوع کارکرد، نمی‌توان به نتایج حاصل شده در منبع [۵] استناد کرد.

جدول ۹ مقایسهٔ نتایج سامانه پیشنهادی با دو روش دیگر.

سامانه	عمق بازیابی	حالت خالص	
		صحت بازیابی (٪)	زمان بازیابی (زمان احتسابی برای سی‌پی‌یو با کد متلب)
روش (۱) [۳۱]	۱	۹۱	۴٫۳ ثانیه
روش (۲) [۵، ۵۷]	۱	۸۴	۱۰٫۲ ثانیه
سامانه پیشنهادی با فاصلهٔ کِی‌ال	۱	۹۵	۳۸۰ میلی‌ثانیه
سامانه پیشنهادی با فاصلهٔ اقلیدسی	۱	۹۱	۵۲۵ میلی‌ثانیه

بررسی نتایج بیان‌شده در جدول ۹ نشان می‌دهد سرعت عملکرد سامانه پیشنهادی در این مقاله بسیار بالاتر از دو روش آزمایش‌شدهٔ دیگر می‌باشد. این در حالی است که صحت بازیابی سامانه پیشنهادی، بیش‌تر یا برابر صحت بازیابی آن دو روش می‌باشد. این امر نشان می‌دهد با به‌کارگیری تشخیص ژانر، ضمن دستیابی به مقدار قابل قبولی از صحت بازیابی، می‌توان سرعت عملکرد سامانه را نیز تا حد زیادی بهبود داد.

۴. نتیجه‌گیری

در این مقاله، یک سامانه بازیابی موسیقی با دریافت یک نمونه علامت پرس‌وجو مبتنی بر تشخیص ژانر پیشنهاد شده است. در این سامانه، تشخیص ژانر به‌عنوان یک عامل کمکی جهت بازیابی مطلوب در زمان مناسب حتی در شرایط نوفه‌ای مورد توجه قرار می‌گیرد. هم‌چنین نشان داده شد از آن‌جا که مشخصهٔ ژانر به خوبی می‌تواند تفاوت نوع‌های مختلف موسیقی را بازگو کند، قادر است تا در پیاده‌سازی یک سامانه بازیابی موسیقی دقیق و سریع کمک شایانی نماید.

بازیابی دقیق و سریع حتی در حضور نوفهٔ پس‌زمینه از جملهٔ خصیصه‌های کلیدی و مطلوب در یک سامانه بازیابی موسیقی جهت تبدیل به یک برنامهٔ کاربردی بر روی گوشی‌های همراه یا یک وب‌سایت جستجوی موسیقی می‌باشد. بنابراین، سامانه بازیابی موسیقی پیشنهادی در این مقاله، می‌تواند سرآغازی برای رسیدن به اهداف فوق باشد.

نتایج سامانه پیشنهادی بازیابی موسیقی نشان داد که استفاده از عامل کمکی تشخیص ژانر علامت پرس‌وجو می‌تواند باعث کاهش فضای جستجو، کاهش هزینهٔ محاسبات و در نتیجه کاهش زمان بازیابی شود. این در حالی است که اگر چه سامانهٔ پیشنهادی بازیابی موسیقی مبتنی بر تشخیص ژانر می‌باشد، اما برخلاف بعضی از برنامه‌های کاربردی و موتورهای جستجوی موسیقی موجود، هیچ نیازی به دانستن یا انتخاب ژانر توسط کاربر نیست. این امر، عملیاتی شدن سامانه پیشنهادی را تسهیل می‌کند. البته، پیشنهاد می‌شود در پژوهش‌های آتی، انواع دیگر اعوجاج‌هایی که می‌توانند در یک سامانه بازیابی موسیقی وجود داشته باشند و باعث افت عملکرد آن شوند نیز مورد بررسی قرار گیرند. عدم تطابق زمانی علامت پرس‌وجو با آهنگ متناظر در دادگان به‌دلیل قطع شدن بخشی از علامت پرس‌وجو یا تند و کند شدن آن، تغییرات کوچک در بسآمد علامت پرس‌وجو به‌دلیل تبدیل قالب[1]، و هم‌چنین تغییرات اندک در نُت‌های علامت پرس‌وجو به‌دلیل بازترکیب[2]، نمونه‌هایی از این اعوجاج‌ها محسوب می‌شوند.

عملکرد سامانه بازیابی موسیقی با استفاده از دادگان جی‌تی‌زان و از طریق دو دستهٔ صدتایی علامت پرس‌وجوی خالص و نوفه‌ای، مورد ارزیابی قرار گرفت. ارزیابی نشان

[1] Format
[2] Re-mix

[DOR: 20.1001.1.23455748.1395.4.1.2.9]

داد نتایج بازیابی در حدود روش جستجوی پایه برای دسته علامت پرس‌وجوی خالص همراه با کاهش قابل توجه در زمان بازیابی، حاصل می‌شود. برای دسته علامت پرس‌وجوی نوفه‌ای، موفقیت بازیابی سامانه پیشنهادی کمی از روش جستجوی پایه کم‌تر است. اما هم‌چنان، نتایج بازیابی با لحاظ کردن زمان بازیابی کوتاه و وجود میزان نوفه بیش از حالات واقعی مطلوب می‌باشد. علاوه بر این‌که فاصلهٔ کِی‌اِل عملکرد بهتری را در حالت خالص نسبت به فاصلهٔ اقلیدسی نشان می‌دهد. البته، عملکرد فاصلهٔ اقلیدسی در حالت نوفه‌ای بهتر می‌باشد.

چنان‌چه در بخش ۳-۵ ذکر شد، دستیابی به میزان صحت بازیابی تا حدود ۹۵٪ برای یک سامانه بازیابی موسیقی مبتنی بر نمونه، کفایت می‌کند و سامانه پیشنهادی در این مقاله تقریباً توانسته است این شرایط را با آزمایش روی یک دادگان خاص محقق سازد. در حالی‌که با کاربرد تشخیص ژانر باعث کاهش قابل توجهی در زمان بازیابی نیز شده است. هم‌چنین، نتایج بلوک اصلی تشخیص ژانر تایید می‌نماید که در تشخیص ژانر، درخت تصمیم دودویی موفق عمل کرده است. علاوه بر این‌که ضرایب کپسترال مِل که برای آموزش درخت تصمیم مورد استفاده قرار گرفته‌اند، به خوبی توانسته‌اند اطلاعات بافتی و طنینی موسیقی‌های مختلف را حتی در شرایط وجود نوفه از یکدیگر تفکیک کنند و موفقیت درخت تصمیم دودویی در تشخیص ژانر و در نتیجه، کسب نتایج بازیابی مطلوب را در پی داشته باشد. به علاوه در بلوک تطبیق- بازیابی، روش ترکیب تصمیم پیشنهادی نیز به خوبی توانسته است آهنگ متناظر با علامت پرس‌وجو را با کم‌ترین هزینهٔ محاسباتی بازیابی کند.

به‌طورکلی، نتایج سامانه پیشنهادی بازیابی موسیقی برای دو دسته علامت پرس‌وجوی خالص و نوفه‌ای، دستیابی به میزان صحت مطلوب در بازیابی را بازگو می‌کند. علاوه بر این‌که بهبود قابل‌توجه زمان بازیابی را در مقایسه با روش جستجوی پایه نشان‌می‌دهد. در واقع، نتایج تایید می‌کند که سامانه بازیابی موسیقی پیشنهادی در این مقاله می‌تواند به عنوان یک راهبُرد مفید در پیاده‌سازی یک سامانه بازیابی موسیقی سریع، دقیق و بلادرنگ به‌صورت یک برنامهٔ کاربردی در گوشی‌های‌همراه یا به‌صورت وب‌سایت جستجوی موسیقی، مورد استفاده قرار گیرد.

۵. فهرست منابع

[1] N. Borjian, E. Kabir, S. Seyedin, E. Masehian, "Music retrieval using Gaussian mixture model," The Fourth International Conference on Acoustics and Vibration, Iran University of Science and Technology, Tehran, Iran, 2014.

[2] N. Borjian, E. Kabir, S. Seyedin, E. Masehian, "Genre recognition-based music retrieval," The 12th Iranian Conference on Intelligent Systems, Higher Education Complex of Bam, 2014.

[3] S. Kiranyaz, "Advanced techniques for content-based management of multimedia databases," PhD Thesis, Tampere University of Technology, Finland, 2005.

[4] A. Meng, P. Ahrendt, J. Larsen, L. K. Hansen, "Temporal feature integration for music genre classification," IEEE Transactions on Audio, Speech and Language Processing, vol. 15, pp. 1654-1664, 2007.

[5] M. Helén, T. Virtanen, "Audio query by example using similarity measures between probability density functions of features," EURASIP Journal on Audio, Speech, and Music Processing, vol. 2010, pp. 1-12, 2010.

[6] T. Dharani, I.L. Aroquiaraj, "A survey on content based image retrieval," International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME 2013) Salem, USA, 2013.

[7] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval," IEEE Transaction on Pattern Analysis and Machine Intelligence vol. 35, no. 12, pp. 2916-2929, 2013.

[8] J.S. Downie, "The International Society of Music Information Retrieval," Available: http://www.ismir.net/, 2000.

[9] M.A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, M. Slaney, "Content-based music information retrieval: Current directions and future challenges," Proceedings of the IEEE, vol. 96, pp. 668-696, 2008.

[10] Z.W. Ras, A. Wieczorkowska, "Advances in Music Information Retrieval," First Edition, Springer-Verlag Berlin Heidelberg, 2010.

[11] M. Schedl, E. Gómez, J. Urbano, "Music information retrieval: Recent developments

and applications," Foundations and Trends in Information Retrieval, vol. 8, pp. 127-261, 2014.

[12] D. Byrd, T. Crawford, "Problems of music information retrieval in the real world," Information Processing and Management, vol. 38, pp. 249-272, 2002.

[13] G. Haus, M .Longari, E. Pollastri, "Score-driven approach to music information retrieval," Journal of the American Society for Information Science and Technology, vol. 55, pp. 1045-1052, 2004.

[14] N.H. Adams, M.A. Bartsch, G.H. Wakefield, "Note segmentation and quantization for music information retrieval," IEEE Transactions on Audio, Speech and Language Processing, vol. 14, pp. 131-141, 2006.

[15] C.N. Silla, A.L. Koerich, C.A.A. Kaestner, "Feature selection in automatic music genre classification," The Tenth IEEE International Symposium on Multimedia (ISM 2008), Berkeley, CA, 2008.

[16] M. Kaminskas, F. Ricci, "Contextual music information retrieval and recommendation: State of the art and challenges," Computer Science Review, vol. 6, pp. 89-119, 2012.

[17] J. Makhoul, F. Kubala, T. Leek, D. Liu, L. Nguyen, R. Schwartz, A. Srivastava, "Speech and language technologies for audio indexing and retrieval," Proceedings of the IEEE, vol. 88, pp. 1338-1353, 2000.

[18] W.-H. Tsai, H.-M. Yu, H.-M .Wang, "Query-by-example technique for retrieving cover versions of popular songs with similar melodies," The Sixth International Conference on Music Information Retrieval, London, UK, 2005.

[19] I.S.H. Suyoto, A.L. Uitdenbogerd, F. Scholer, "Effective retrieval of polyphonic audio with polyphonic symbolic queries," The MIR '07 Proceedings of the International Workshop on Multimedia Information Retrieval, 2007.

[20] H.-M. Yu, W.-H. Tsai, H.-M. Wang, "A query-by-singing system for retrieving karaoke music," IEEE Transactions on Multimedia, vol. 10, pp. 1626-1637, 2008.

[21] W.-H. Tsai, Y.-M. Tu, C.-H. Ma, "An FFT-based fast melody comparison method for query-by-singing/humming systems," Pattern Recognition Letters, vol. 33, pp. 2285-2291, 2012.

[22] E. Unal, E. Chew, P.G. Georgiou, S.S. Narayanan, "Challenging uncertainty in query by humming systems: A fingerprinting approach," IEEE Transactions on Audio, Speech and Language Processing, vol. 16, pp. 13, 2008.

[23] S. Doraisamy, S. Ruger, "Robust polyphonic music retrieval with n-grams," Journal of Intelligent Information Systems, vol. 21, pp. 53-70, 2003.

[24] J. Allali, P. Ferraro, P. Hanna, C. Iliopoulos, M. Robine, "Toward a general framework for polyphonic comparison," Fundamenta Informaticae, vol. 97, pp. 331-334, 2009.

[25] J. Salamon, E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," IEEE Transactions on Audio, Speech and Language Processing, vol. 20, pp. 1759-1770, 2012.

[26] R.P. Paiva, T. Mendes, A. Cardoso, "A methodology for detection of melody in polyphonic musical signals," The 116th Convention, Audio Engineering Society, Berlin, Germany, 2004.

[27] H.-M. Yu ,W.-H. Tsai, H.-M. Wang, "A query-by-singing technique for retrieving polyphonic objects of popular music," Information Retrieval Technology Lecture Notes in Computer Science, vol. 3689, pp. 439-453, 2005.

[28] M. Kle´c, D. Kor˘zinek, "Unsupervised feature pre-training of the scattering wavelet transform for musical genre recognition," Procedia Technology, vol. 18, pp. 133-139, 2014.

[29] M. Banitalebi-Dehkordi, A. Banitalebi-Dehkordi, "Music genre classification using spectral analysis and sparse representation of the signals," Journal of Signal Processing Systems, vol. 74, pp. 273-280, 2014.

[30] R. Mayer, R. Neumayer, A. Rauber, "Combination of audio and lyrics features for genre classification in digital audio collections," The MM '08 Proceedings of the Sixteenth ACM International Conference on Multimedia, New York, USA, 2008.

[31] C. Barton, P. Inghelbrecht, A. Wang, D. Mukherjee, Shazam Company, Available: http://www.shazam.com/company, 1999.

[32] G. Tzanetakis, P. Cook, "Musical genre classification of audio signals," IEEE Transactions on Speech and Audio Processing, vol. 10, pp. 293-302, 2002.

[33] R. Kaye, "The Open Music Encyclopedia," Available: https://musicbrainz.org/, 2005.

[34] A. Schröder, M. Keith, Free database Available: http://www.freedb.org.

[35] K. Mohajer, M. Emami, J. Hom, K. McMahon, T. Stonehocker, C. Lucanegro, K. Mohajer, A. Arbabi, and F. Shakeri. www.soundhound.com, 2010.

[36] J. Born, Neuros, Available: www.neurostechnology.com, 2001.

[37] A.L.-C. Wang, "An industrial strength audio search algorithm," Proceedings of the Fourth International Conference on Music Information Retrieval (ISMIR 2003), Baltimore, MD, 2003.

[38] W.-H. Tsai, H.-M. Yu, and H.-M. Wang, "Query-by-example technique for retrieving cover versions of popular songs with similar melodies," Proceedings of the Sixth International Conference on Music Information Retrieval, London, 2005.

[39] K. Itoyama, M. Goto, K. Komatani, T. Ogata, H.G. Okuno, "Query-by-example music information retrieval by score-informed source separation and remixing technologies," EURASIP Journal on Advances in Signal Processing, pp. 1-14, 2010.

[40] G. Tzanetakis, "GTZAN genre collection," Available: http://marsyas.info/downloads/datasets.html

[41] L.R. Rabiner, B.H. Juang, "Fundamental of Speech Recognition Prentice," First Edition, Prentice Hall, 1993.

[42] F. Camastra, A. Vinciarelli, "Machine Learning for Audio, Image and Video Analysis: Theory and Applications," London, Springer, 2008.

[43] Z.H. Tan, B. Lindberg, "Automatic Speech Recognition on Mobile Devices and Over Communication Networks," London, Springer-Verlag, 2008.

[44] U.S. Tiwary, T.J. Siddiqui, "Speech, Image, and Language Processing for Human Computer Interaction: Multi-modal Advancements," USA, IGI Global, 2012.

[45] J. Shen, J. Shepherd, A.H.H. Ngu, "Towards effective content-based music retrieval with multiple acoustic feature combination," IEEE Transactions on Multimedia, vol. 8, pp. 1179-1189, 2006.

[46] L. Breiman, J. Friedman, R. Olshen, C. Stone, Classification and Regression Trees: Chapman and Hall, 1984.

[47] J.R. Quinlan, "Induction of decision trees," Machine Learning, vol. 1, no. 1, pp. 81-106, 1986.

[48] Y. Riblet, "Decision Tree Learning," Available: http://en.wikipedia.org/wiki/Decision_tree_learning, 2007.

[49] I. Narsky, F.C. Porter, "Statistical Analysis Techniques in Particle Physics, Fits, Density Estimation and Supervised Learning," Wiley, 2013.

[50] J. Goldberger, S. Gordon, H. Greenspan, "An efficient image similarity measure based on approximations of KL-divergence between two Gaussian mixtures," Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, 2003.

[51] E.G.I. Termens, "Audio content processing for automatic music genre classification: Descriptors, databases, and classifiers," PhD Thesis, Department of Information and Communication Technologies, University Pompeu Fabra, Barcelona, 2009.

[52] M.B. Dehkordi, "Music genre classification using spectral analysis and sparse representation of the signals," Journal of Signal Processing Systems, vol. 74, pp. 8, 2014.

[53] M. Helen, "Similarity measures for content-based audio retrieval," PhD Thesis, Tietotalo Building, Tampere University of Technology, Finland, 2009.

[54] M. Gowan, Available: http://www.techhive.com/, 2011.

[55] I. Cox, M. Miller, J. Bloom, J. Fridrich, T. Kalker, "Digital Watermarking and Steganography," Second Edition, Morgan Kaufmann, 2007.

[56] J. Haitsma, A. Kalker, "A highly robust audio fingerprinting system with an efficient search strategy," Journal of New Music Research, vol. 32, pp. 211-222, 2003.

[57] B. Thoshkahna, K. Ramakrishnan, "Arminion:a query by example system for audio retrieval," Proceedings of Computer Music Modelling and Retrieval, pp. 1-9, 2005.

Journal of Acoustical Engineering Society of Iran, Vol. 4, No. 1, 2016

Query-by-example music retrieval using genre recognition to speed up the performance

N. Borjian*[1], E. Kabir[1], S. Seyedin[2], E. Masehian[3]

1. Department of Electrical and Computer Engineering, Tarbiat Modares Univ.
2. Department of Electrical Engineering, Amirkabir University of Technology
3. Faculty of Engineering, Tarbiat Modares Univ.

Abstract

The goal of a query-by-example music information retrieval system is retrieval of the target song corresponding to user-provided example from a particular dataset. The example can be a few second piece recorded from any music source such as TV or even a noisy environment e.g. gym. In this paper, a query-by-example system for music retrieval using genre recognition is proposed whose goal is to show the effect of genre recognition to achieve the accurate and rapid performance in such systems even in presence the background noise. This system includes two basic blocks: genre recognition and matching-retrieval. A binary decision tree performs the genre recognition and matching-retrieval uses two Euclidean and Kullback-Leibler (KL) distances along with a score level based decision fusion. The proposed system is evaluated on the well-known GTZAN dataset (prepared by George Tzanetakis) and by two random groups of pure and noisy queries. The results show the accuracy of 97% and 86% for two pure and noisy query groups, respectively, in retrieval time of 525 ms with Euclidean distance. These values are 97% and 82% in retrieval time of 380 ms with KL distance.

Keywords: Music information retrieval, Query by example, Genre recognition, Decision fusion, Noise.

pp. 1-20 (In Persian)

[DOR: 20.1001.1.23455748.1395.4.1.2.9]

* Corresponding author E-mail: nastaran.borjian@modares.ir