طبقه‌بندی دستگاه‌های موسیقی سنتی ایرانی با استفاده از ویژگی‌های مبتنی بر فرکانس گام

مسعود گراوانچی‌زاده[1*]، پریسا مبشری[2] و هادی جمشیدی اوانکی[3]

[1و3]دانشکده مهندسی برق و کامپیوتر دانشگاه تبریز، تبریز، ایران

[2]دانشکده مهندسی برق و کامپیوتر، دانشگاه آزاد اسلامی واحد علوم و تحقیقات آذربایجان شرقی، تبریز، ایران

چکیده

در این مقاله، سامانه جدیدی با استفاده از ویژگی‌های مبتنی بر فرکانس گام (Pitch) جهت طبقه‌بندی دستگاه‌های موسیقی سنتی ایرانی ارائه شده است. موسیقی سنتی ایرانی از هفت دستگاه اصلی شامل چهارگاه، همایون، ماهور، سه‌گاه، شور، نوا و راست‌پنجگاه تشکیل می‌شود. در این الگوریتم، ویژگی‌های ضرایب لاگرانژ لگاریتم فرکانس گام (LCPL)، دسته‌های شباهت فازی نوع دو (FSST2) و نیز ترکیب آنها جهت طبقه‌بندی سیگنال‌های موسیقی به‌کار برده می‌شود و از ماشین‌بردار پشتیبان چندرده‌ای به‌منظور طبقه‌بندی دستگاه‌های موسیقی ایرانی استفاده می‌شود. دادگان به‌کاررفته در سامانه جداساز ارائه‌شده، شامل قطعاتی از تک‌نوازی‌های تار علیزاده، استاد مطرح موسیقی ایرانی، است. عملکرد الگوریتم‌های طبقه‌بندی پایه و پیشنهادی توسط معیارهای ارزیابیAccuracy، Recall، Precision، F-measure و MCC صورت می‌گیرد. نتایج به‌دست‌آمده نشان می‌دهند که الگوریتم پیشنهادی نسبت به روش‌های پایه عملکرد بهتری را بر حسب معیارهای مختلف طبقه‌بندی در بر دارد.

واژگان کلیدی: طبقه‌بندی موسیقی، فرکانس گام، منطق فازی نوع دو، ماشین‌بردار پشتیبان چندکلاسی، دستگاه‌های موسیقی ایرانی.

Classification of Iranian Traditional Music Dastgahs Using Features Based on Pitch Frequency

Masoud Geravanchizadeh[*1], Parisa Mobasheri[2] & Hadi Jamshidi Avanaki[3]

[1,3]Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran

[2]Faculty of Electrical and Computer Engineering Islamic Azad University, Tabriz, Iran

Abstract

The Iranian traditional music is composed of seven majors Dastgahs: Chahargah, Homayoun, Mahour, Segah, Shour, Nava, and Rast-Panjgah. In this paper, a new algorithm for the classification of the Iranian traditional music Dastgahs based on pitch frequency is proposed. In this algorithm, the features of Lagrange coefficients of pitch logarithm (LCPL), Fuzzy similarity sets type 2 (FSST2), and their combination are used as the representation of music signals which are fed into the multi-class support vector machine (MSVM) as the classifier.

The features of LCPL and FSST2 are obtained by applying some modifications on the pitch frequency of the desired music. To compute LCPL, first, the values of pitch frequency are extracted by the PRAAT algorithm. Then, after the applying a logarithmic operation, the tracks of pitch frequency are partitioned into smaller segments. The method of feature extraction is based on detecting the trough or valley points of the pitch tracks. In the following, the coordinates of trough points (i.e., the index of pitch frequency and the logarithmic value of the frequency) are considered as each segment boundaries. In the next step, the track between the two boundaries of each segment is approximated by a 6th order Lagrange polynomial and the computed polynomial coefficients are considered as a 6-dimensional feature vector.

The first step in extracting the FSST2 feature is to compute the pitch frequencies of the input signal by the PRAAT algorithm. The second step involves the classification of music notes. Then, the subtractive

* نویسندهٔ عهده‌دار مکاتبات

* Corresponding author

• تاریخ ارسال مقاله: ۱۳۹۹/۴/۲۰ • تاریخ پذیرش: ۱۴۰۰/۹/۲۰ • تاریخ انتشار: ۱۴۰۱/۱۰/۳ • نوع مطالعه: پژوهشی

clustering method is used to eliminate the incorrectly estimated pitch frequencies of the previous step. Next, the process of folding notes (i.e., transferring the extracted pitch frequencies into the reference octave band of 220-440 Hz) is performed followed by translating the frequency points to the cents with respect to 220 Hz. After folding notes in one octave, the Mahalanobis distance is applied to recognize which point on the reference octave corresponds to each musical note. These same procedures are conducted for the information pattern (theoretical data) of each Dastgah. In the final step, the folded frequency points of the unknown input signal and the information pattern of all Dastgahs are transferred to the Fuzzy logicType-2 domain and compared to determine a similarity measure which is considered as the extracted feature.
The dataset used in the proposed classification algorithm contains the excerpts from solo performances with Tar played by Alizadeh, the well-known Iranian music master. The performances of the baselines and proposed classification algorithms are evaluated by the measures of Accuracy, Recall, Precision, F-measure, and MCC. The results show that the proposed algorithm has a better performance as compared with the baseline methods in terms of different classification criteria.

Keywords: Music Classification, Pitch Frequency, Fuzzy Logic Type 2, MSVM, Iranian Music Dastgahs.

۱- مقدمه

الگوریتم‌های طبقه‌بندی موسیقی در بسیاری از زمینه‌های مرتبط با پردازش سیگنال‌های موسیقی رشد کرده است، اما بین طبقه‌بندی سیگنال‌های موسیقی غربی و غیرغربی، به‌ویژه موسیقی خاورمیانه، فاصلهٔ قابل‌توجهی وجود دارد. یکی از مهم‌ترین چالش‌های انجمن بین‌المللی بازیابی اطلاعات موسیقی[۱] (ISMIR)، گسترش پژوهش‌های پردازش سیگنال‌های موسیقی به سمت موسیقی غیرغربی است. عمدهٔ پژوهش‌هایی که در طبقه‌بندی سیگنال‌های صوتی صورت گرفته است، تشخیص و شناسایی سیگنال‌های صحبت و جداسازی آن‌ها از موسیقی بوده و در زمینهٔ سیگنال‌های موسیقی ایرانی به‌تنهایی، کار چندانی انجام نشده است. چالش اصلی که پژوهش‌گران با آن روبه‌رو هستند، یافتن ویژگی‌هایی است که بتوان به‌وسیلهٔ آن‌ها سیگنال‌های موسیقی را به‌دقت در طبقات مختلف طبقه‌بندی کرد. روش‌های استخراج ویژگی سیگنال‌های موسیقی بسیار محدود است. ویژگی‌های متداول مورد استفاده در تحلیل سیگنال‌های موسیقی عبارتند از: فرکانس گام[۲]، ضرایب ناموزونی[۳]، ضرایب طیفی فرکانس مل[۴] (MFCCs)، نرخ گذر از صفر[۵]، نقطه Roll-Off طیفی، انرژی مدولاسیون چهار هرتز و غیره [1]. در موسیقی ایرانی، جداسازی را می‌توان درخصوص نوع دستگاه، مقام و ردیف انجام داد. در مورد پژوهش‌های صورت‌گرفته در این زمینه می‌توان به پژوهش انجام‌شده توسط *سجاد عبدلی* [2] اشاره کرد. سجاد عبدلی سامانه‌ای را جهت شناسایی دستگاه‌های موسیقی سنتی ایرانی مبتنی بر دسته‌های فازی وقفه‌ایی نوع دو[۶] (IT2FSs) معرفی کرده است. در این سامانه، پس از استخراج ویژگی، هر نت اجراشده به‌عنوان یک IT2FS در نظر گرفته می‌شود؛ بنابراین، هر قطعهٔ موسیقی از چندین IT2FS تشکیل شده است. بیشترین شباهت بین این مجموعه‌ها و ساختار دستگاه‌ها[۷]، که آن نیز از چندین IT2FS تشکیل شده است، دستگاه مورد نظر را شناسایی می‌کند.

محمود عباسی لایق و همکارانش [1]، الگوریتمی برای جداسازی ردیف‌های میرزا عبدالله از طریق جداساز ماشین بردار پشتیبان[۸] (SVM) ارائه کرده‌اند. در این پژوهش، ردیف‌های استاد سه تار، میرزا عبدالله، در هفت مجموعه بزرگ (دستگاه‌های شور، ماهور، همایون، سه‌گاه، چهارگاه، راست‌پنجگاه و نوا) جداسازی شده‌اند. در این سامانه، ابتدا، یک سری پیش‌پردازش از جمله تقسیم‌کردن سیگنال به چندین بخش، هنجارسازی[۹] هر بخش، کاهش نرخ نمونه‌برداری، و درنهایت، پنجره‌بندی سیگنال صورت می‌پذیرد. در مرحلهٔ بعد که استخراج ویژگی است، چهار ویژگی فرکانس گام، ضرایب ناموزونی، MFCC و مرکز طیفی[۱۰] (SC) از سیگنال موسیقی استخراج می‌شوند. پس از استخراج ویژگی‌های بالا، از طریق ماشین بردار پشتیبان، جداسازی ردیف‌های میرزا عبدالله صورت می‌پذیرد.

دارابی و همکارانش [3]، از روش‌های بازشناسی الگو[۱۱] برای شناسایی دستگاه و مقام موسیقی سنتی ایرانی استفاده کرده‌اند. در این، یک سری اندازه‌گیری‌های آماری تعریف می‌شود که خصوصیات ملودیک سیگنال موسیقی را بیان می‌کنند؛ سپس، از یک الگوی استاندارد ادراکی، جهت شناسایی الگو استفاده شده است. در این روش، هر

[1] International Society for Music Information Retrieval (ISMIR)
[2] Pitch
[3] Inharmonicity
[4] Mel-Frequency Cepstral Coefficients (MFCCs)
[5] Zero-Crossing Rate (ZCR)
[6] Interval Type 2 Fuzzy Sets (IT2FSs)
[7] Prototypes
[8] Support Vector Machine (SVM)
[9] Normalization
[10] Spectral Centroid (SC)
[11] Pattern Recognition

دستگاه با یک آرایهٔ عددی مدل می‌شود که این آرایه نشان‌گر نسبت فواصل فرکانسی نت‌های به کار رفته در آن دستگاه نسبت به نت پایهٔ آن است.

درخصوص موسیقی غیرایرانی، می‌توان جداسازی را از نظر نوع سبک موسیقی انجام داد. *Ghosal و همکارانش* [4] در مقالهٔ خود از یک روند ترتیبی جهت طبقه‌بندی اطلاعات موسیقی استفاده کرده‌اند. آنها به‌جای اینکه از تعداد زیادی ویژگی استفاده کنند، فقط از MFCC و تغییرات آن استفاده کرده‌اند. در این پژوهش، موسیقی به دو دستهٔ موسیقی باکلام و موسیقی بی‌کلام طبقه‌بندی می‌شود. در هر یک از این دو دسته نیز موسیقی بر اساس سبک و نوع ساز طبقه‌بندی می‌شود. با استفاده از پردازش دومرحله‌ای، سیگنال‌های موسیقی بدون‌کلام، بر اساس یکی از چهار نوع String، Woodwind، Percussion و یا Keyboard شناسایی می‌شوند. برای طبقه‌بندی بخش موسیقی باکلام در مرحلهٔ نخست، با به‌دست‌آوردن الگوهای MFCC سیگنال تجزیه‌شدهٔ باند میانی، سیگنال‌ها به دو دستهٔ کلاسیک و غیرکلاسیک تقسیم می‌شوند. در مرحلهٔ دوم، طبقه‌بندی صوت‌های معروف در سبک‌های متفاوت، مثل پاپ، جاز، Bhangra (یک سبک هندی)، بر اساس الگوی تغییر دامنه انجام می‌شود. در همهٔ مراحل، از الگوریتم RANSAC[1] به‌عنوان جداساز استفاده شده است.

Wülfing و *Riedmiller* [5] از روش آموزش بدون ناظر ویژگی‌ها، جهت پیش‌گویی سبک موسیقی استفاده کرده‌اند. آنها روشی را که درهمین‌اواخر با موفقیت برای تشخیص اشیا در تصاویر به‌کار برده شده است، مورد ارزیابی قرار داده‌اند. ابتدا، تکه‌هایی محلی[2] از تبدیل زمان-فرکانسی سیگنال صوتی استخراج می‌شوند؛ سپس، این تکه‌ها پیش‌پردازش شده و برای آموزش بدون ناظر مورد استفاده قرار می‌گیرند. برای آموزش، از روش خوشه‌بندی Bootstrapped K-means و یا انتخاب ویژگی به‌طور تصادفی استفاده می‌شود. همچنین، ویژگی‌ها با روش کانولوشن استخراج شده و ماشین بردار پشتیبان خطی نیز جهت طبقه‌بندی آموزش داده می‌شود.

درنهایت، عملکرد سامانه جداساز با دادگان GTZAN [6] ارزیابی می‌شود. نشان داده شده که استفاده از این تکنیک نسبت به روش‌های قبلی سریع‌تر و آسان‌تر است.

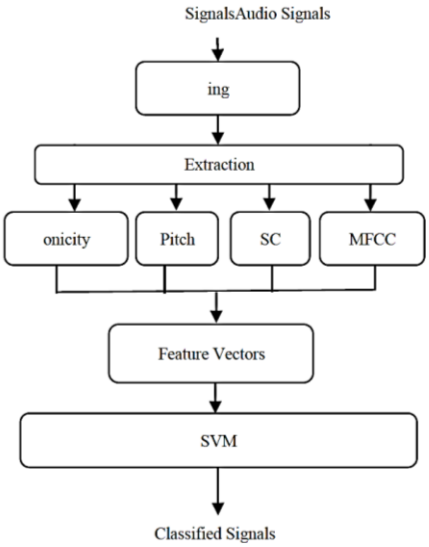
Tzanetakis و همکارانش [7] روشی به‌منظور طبقه‌بندی خودکار سبک‌های موسیقی بر اساس ویژگی‌هایی که ساختار ریتمیک موسیقی را بیان می‌کنند، ارائه کرده‌اند. در این پژوهش، در دو سطح استخراج ویژگی صورت گرفته است: ۱- ویژگی‌های سطحی موسیقی[3]، ۲- ویژگی‌های ریتمیک موسیقی. ویژگی‌های سطحی موسیقی به ویژگی‌هایی می‌گویند که مربوط به متن موسیقی، طنین و آلات به‌کاررفته در آن است. بدین منظور از بردار ویژگی‌ها که شامل ویژگی‌های میانگین مرکز ثقل[4]، میانگین نقطهRolloff[5]، میانگین شار[6]، میانگین نقطهٔ عبور از صفر[7]، مرکز ثقل انحراف معیار[8]، Rolloff انحراف معیار[9]، شار انحراف معیار[10] و عبور از صفر انحراف معیار[11] است، استفاده شده است. محاسبه ویژگی‌های ریتمیک به‌وسیلهٔ تبدیل موجک[12] انجام می‌شود. درنهایت، جداساز گوسین برای طبقه‌بندی به‌کار می‌رود.

Song و همکارانش [8] در سال ۲۰۱۲ موسیقی را از نظر احساسی طبقه‌بندی کرده‌اند. در کل، موسیقی از نظر احساسی به انواع شاد، غمگین، دارای خشم، آرامش‌بخش و غیره تقسیم می‌شود. این که موسیقی چه احساسی را القا می‌کند، بستگی به ویژگی‌هایی از جمله ریتم، طنین، متن آهنگ[13] و طیف دارد. در ابتدا، ویژگی‌های موسیقی را در چهار بُعد پویایی، طیفی، ریتمیک و هارمونیک استخراج کرده و از ماشین بردار پشتیبان جهت جداسازی استفاده کرده‌اند. ویژگی‌های به‌کاررفته در این پژوهش که در چهار بُعد بیان‌شده استخراج می‌شوند، شامل ۵۵ ویژگی بوده و جهت محاسبه آنها از جعبه‌ابزار MIR Toolbox [9] بهره برده شده است.

جدیدترین جداسازها که در سال‌های اخیر در طبقه‌بندی موسیقی به کار گرفته شده‌اند، شبکه عصبی کانولوشن[14] (CNN) و شبکه عصبی کانولوشن تکرارشونده[15] (CRNN) بوده [20, 21, 22] که نسبت به شبکه عصبی مصنوعی و سایر جداسازها نتایج بهتری را نشان داده‌اند. در این پژوهش‌ها از ویژگی‌های زمان-فرکانسی استفاده شده که به‌وسیلهٔ آنها سبک‌های مختلف موسیقی را طبقه‌بندی کرده‌اند.

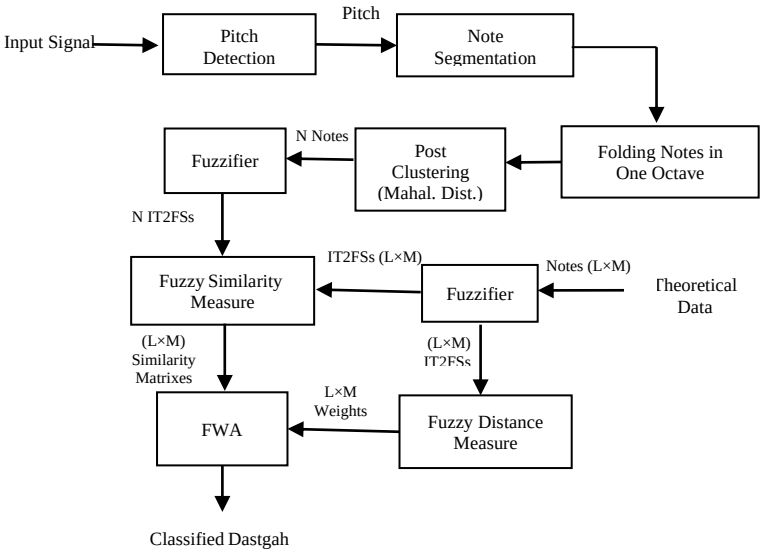
[1] Random Sample Consensus (RANSAC)
[2] Local Patches

[3] Musical Surface Features
[4] Mean-Centroid
[5] Mean-Rolloff
[6] Mean-Flux
[7] Mean-Zero Crossing
[8] Std-Centroid
[9] Std-Rolloff
[10] Std-Flux
[11] Std-Zero Crossing
[12] Wavelet
[13] Lyric
[14] Convolutional Neural Network (CNN)
[15] Convolutional Recurrent Neural Network (CRNN)



(شکل-۱): نمودار جعبه‌ای سامانه عباسی لایق جهت طبقه‌بندی دستگاه‌های موسیقی سنتی ایرانی [1]. در این الگوریتم، ابتدا، چهار ویژگی MFCC، SC، Pitch و Inharmonicity استخراج شده‌اند؛ سپس، با استفاده از جداساز SVM عمل جداسازی صورت می‌گیرد.

(Figure-1): The block diagram of the Abbassi Layegh's system for the classification of Iranian traditional music Dastgahs [1]. In this algorithm, first, four features MFCC, SC, Pitch, and Inharmonicity are extracted. Then, SVM is used for classification.



(شکل-۲): نمودار جعبه‌ای سامانه عبدلی جهت دسته‌بندی دستگاه‌های موسیقی سنتی ایرانی [2]. در این الگوریتم، ابتدا، فرکانس پایه سیگنال ورودی استخراج می‌شود؛ سپس، پس از اعمال فرآیند فازی‌سازی بر روی سیگنال‌های ورودی و سیگنال‌های اطلاعات تئوری، عمل دسته‌بندی با استفاده از الگوریتم منطق فازی نوع ۲ صورت می‌پذیرد.

(Figure-2): The block diagram of the Abdoli's system for the classification of Iranian traditional music Dastgahs [2]. In this algorithm, first, the pitch frequency of input signal is extracted. Then, after applying the fuzzy type 2 process on input signals and the theoretical signals, the fuzzy logic type 2 algorithm is used for classification.

در سال ۲۰۱۶، بیگ‌زاده و گروهش [23] از ۲۰ قلهٔ اصلی طیف موسیقی به‌عنوان ویژگی و از شبکه عصبی مصنوعی به‌عنوان جداساز استفاده کرده و دستگاه‌های موسیقی سنتی ایرانی را طبقه‌بندی کرده‌اند.

حیدری [24] در رسالهٔ دکترای خود که در سال ۲۰۱۶ در زمینهٔ جداسازی دستگاه‌های موسیقی سنتی ایران ارائه کرده است، پس از بررسی ویژگی‌های متعددی که در رابطه با طبقه‌بندی انواع موسیقی تاکنون به کار رفته، از ویژگی‌های هیستوگرام فرکانس گام، میانگین طیفی، میانگین Chroma استفاده کرده و مدل ترکیبی گوسین[1] (GMM) را برای جداسازی به‌کار برده است.

حبیبی اقدم و همایون‌پور [25] در سال ۲۰۱۰ با استفاده از ویژگی‌های میان‌مدت و کوتاه‌مدت توانستند شناسایی خودکار هشت سبک موسیقی را انجام دهند.

[1] Gaussian Mixture Model (GMM)

ساختار مقالهٔ حاضر بدین‌ترتیب است: ابتدا، در بخش دو، به دو سامانه جداسازی دستگاه‌های موسیقی سنتیِ ایرانیِ طراحی‌شده توسط محمود عباسی لایق [1] و سجاد عبدلی [2] می‌پردازیم. در بخش سه، سامانه پیشنهادی جهت جداسازی دستگاه‌های موسیقی سنتی ایرانی مطرح می‌شود. گزارش نتایج حاصل از سامانه پیشنهادی و سامانه‌های مبنا در بخش چهار و در انتها، مقایسهٔ نتایج سامانه پیشنهادی و سامانه‌های مبنا و نتیجه‌گیری در بخش پنج آمده است.

۲- سامانه‌های پایه جهت طبقه‌بندی دستگاه‌های موسیقی سنتی ایرانی

۲-۱- سامانه عباسی لایق

محمود عباسی لایق [1] سامانه‌ای جهت طبقه‌بندی دستگاه‌های موسیقی سنتی ایرانی از طریق جداساز ماشین بردار پشتیبان ارائه کرده است. در این پژوهش، یک سامانه شناسایی دستگاه مبتنی بر چهار ویژگی فرکانس گام، ضرایب ناموزونی، MFCC و SC معرفی شده است. پس از استخراج ویژگی‌ها، با استفاده از ماشین بردار پشتیبان چندرده‌ای[1] (MSVM) [10, 11] دستگاه مورد نظر شناسایی می‌شود. شکل (۱) نمودار جعبه‌ای این سامانه را نشان می‌دهد. در این نمودار جعبه‌ای، نخستین مرحله پس از ورود سیگنال، عملیات پیش‌پردازش است. این عملیات شامل هنجارسازی سیگنال ورودی، کاهش نرخ نمونه‌برداری و پنجره‌بندی سیگنال است. مرحلهٔ بعد، استخراج ویژگی است که مهم‌ترین قسمت سامانه جداسازی است. در این مرحله چهار ویژگی فرکانس پایه، ضرایب ناموزونی، MFCC و SC از سیگنال موسیقی استخراج می‌شوند. تک‌تک ویژگی‌های یادشده در ادامه به‌تفصیل آمده است.

۲-۱-۱- استخراج ویژگی فرکانس گام

فرکانس گام یک ویژگی ادراکی از اصوات متناوب یا به‌طورتقریبی متناوب طیفی است که شامل هارمونیک‌های یک فرکانس است. تغییرات فرکانس گام نشان‌دهنده زیر و بمی صدا است. فرکانس گام، فرکانس اصلی یک صوت را نشان می‌دهد که به‌همراه قابلیت یا کیفیت صدا[2] و بلندی صدا، سه مشخصه اصلی سیگنال‌های صوتی هستند. مطابق با موسسه استانداردهای ملی آمریکا[3] (ANSI)، فرکانس گام یک مشخصه آکوستیکی صوت است که در آن صداها را می‌توان به‌صورت تغییرات گام از پایین تا بالا در نظر گرفت. درواقع فرکانس گام یک حس درونی است که در آن شنونده اصوات دریافتی را متناسب با موقعیت گام‌های موسیقی بر اساس فرکانس نوسان تعیین می‌کند. این فرکانس به میزان کمتری به سطح فشارصوتی (بلندی، حجم) آهنگ یا صدا، به‌خصوص در فرکانس‌های زیر هزار هرتز و بالای دوهزار هرتز وابسته است. وقتی فشار صوت افزایش می‌یابد گام صداهای پایین‌تر، کمتر می‌شود. وقتی صداها بلندتر می‌شوند گام اصوات بالاتر نیز بیشتر می‌شود.

ضرایب ناموزونی

در موسیقی، ناموزونی یا ناهماهنگی فرکانسی اندازه‌ای است که فرکانس‌های اصوات فرعی از ضرایب کلی فرکانس اصلی فاصله می‌گیرد. از لحاظ صداشناسی، یک نت دارای یک گام مشخص و انواع اصوات فرعی اضافی است. بسیاری از سازهای کوبه‌ای مثل سنج و طبل هندی اصوات ناموزون و پیچیده‌ای تولید می‌کنند؛ اما در سازهای زهی مثل پیانو، ویولون، تار و سه‌تار اصوات فرعی بسیار نزدیک به هم است و بعضی مواقع به‌طوردقیق ضرایبی از فرکانس اصلی هستند. هرگونه انحرافی از این سری هارمونیک، ناموزونی نامیده می‌شود. هر چقدر سیم‌های اداوات موسیقی حالت الاستیکی کمتری داشته باشند، به عبارتی، سیم‌ها کوتاه‌تر و ضخیم‌تر باشند، مقدار ناموزونی بیشتر خواهد بود.

MFCC

MFCC یک نمایش فشرده طبیعی سیگنال است. در این ضریب معیار Mel به‌کار می‌رود که مقیاسی از گام‌های فرکانسی است که از یکدیگر فاصله یکسان دارند. فرکانس در مقیاس هرتز توسط معادله زیر به حوزه کاری Mel تبدیل می‌شود:

SC

SC یا مرکز طیفی، نشان‌دهندهٔ نقطه تعادل توزیع توان طیفی در یک فریم است. بسیاری از انواع سیگنال‌های موسیقی شامل صداهای کوبه‌ای هستند که با در نظر گرفتن نوفه در فرکانس‌های بالا، میانگین طیف را به سمت بالاتر می‌کشانند. علاوه‌بر آن، انرژی‌های تحریک برای

[1] Multi-class Support Vector Machine (MSVM)
[2] Timbre
[3] American National Standards Institute (ANSI)

سیگنال‌های موسیقی بالاتر از سیگنال‌های صحبت است و در آن فرکانس گام در محدوده به‌نسبه پایینی می‌ماند. این معیار برای سیگنال‌های صحبت با کلام و بی‌کلام نتایج مختلفی می‌دهد. مرکز طیفی برای یک فریم، که در یک زمان مشخص رخ می‌دهد، به‌صورت زیر محاسبه می‌شود:

$$SC = \sum_{k=1}^{n-1} \frac{kx(k)}{x(k)}, \qquad (۲)$$

که در آن k یک باند فرکانس کوچک در داخل طیف اندازه‌گیری شده است. $x(k)$ دامنه سیگنال در باند فرکانس معادل بوده و n طول تبدیل فوریه گسسته[1] (DFT) است. فرمول استاندارد برای میانگین مرکز طیفی عبارت است از:

$$c = \sum_{i=1}^{n} \frac{c_i}{i}, \qquad (۳)$$

که در آن c_i مرکز یک فریم طیف است و i تعداد فریم‌ها برای هر صوت است. یک فریم طیف شامل تعدادی از نمونه‌ها است که برابر با اندازه FFT [2] است.

مقدار SC یک فریم، به‌صورت میانگین وزنی فرکانس با دامنه تقسیم بر مجموع دامنه‌ها است که به‌صورت زیر تعریف می‌شود:

$$c_i = \frac{\sum_{i=1}^{n} f_i a_i}{\sum_{i=1}^{n} a_i}, \qquad (۴)$$

که در آن i تعداد فریم‌ها است. در این رابطه، مرکز این فرکانس را برای فریم داده‌شده پیدا می‌کند؛ سپس نزدیک‌ترین بلوک طیف را برای آن فرکانس می‌یابد. در این قسمت، میزان دقت به اندازه فریم بستگی دارد. باید توجه کرد که SC با فرکانس اصلی اشتباه گرفته نشود. هر چه SC بیشتر باشد، صدا واضح‌تر است.

پس از استخراج ویژگی‌های بالا، طبقه‌بندی از طریق ماشین بردار پشتیبان چندرده‌ای صورت می‌پذیرد.

۲-۱-۲- سامانه جداساز MSVM

ماشین بردار پشتیبان یکی از روش‌های یادگیری با نظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند [10]. این جداساز چندرده‌ای درواقع به‌وسیلهٔ ترکیب ماشین بردارهای پشیبان دورده‌ای حاصل شده است. هر مسئله جداسازی چندرده‌ای، ابتدا، با یک جداکننده دودویی حل، سپس، خروجی جداکننده‌های دودویی با هم ترکیب شده و به این ترتیب، مسئله چندرده‌ای حل می‌شود. این روش طبقه‌بندی، ماشین بردار پشتیبان چندرده‌ای (MSVM) نام‌گذاری شده است [11].

۲-۲- سامانه سجاد عبدلی

سجاد عبدلی [2] یک سامانه شناسایی دستگاه‌های موسیقی سنتی ایرانی مبتنی بر دسته‌های فازی وقفه‌ایی نوع دو (IT2FSs) معرفی کرده است. در این سامانه، هر نت اجرایی به‌عنوان یک IT2FS در نظر گرفته می‌شود. بنابراین، هر قطعه موسیقی از چندین IT2FS تشکیل شده است. بیشترین شباهت بین این دسته‌ها و ساختار دستگاه‌ها که آن نیز از چندین IT2FS تشکیل شده، دستگاه موردنظر را شناسایی می‌کند. شکل (۲) نمودار جعبه‌ای این سامانه را نشان می‌دهد. در این نمودار جعبه‌ای، نخستین مرحله، تعیین فرکانس پایهٔ ورودی است. برای این منظور می‌توان از نرم‌افزار PRAAT [12] استفاده کرد. در مرحلهٔ دوم، طبقه‌بندی نت صورت می‌گیرد. در ابتدا، باید مشخص شود چه نت‌هایی در طول اجرای قطعه نواخته شده است. علاوه‌بر آن، باید فرکانس‌های پایه‌ای را که در مرحلهٔ قبل ممکن است اشتباه تخمین زده شده باشند، حذف کرد. در این مقاله، از روش خوشه‌بندی کاهشی[3] [13] استفاده شده است. برای سهولت محاسبات، بهتر است تمام مراکز استخراج‌شده در مرحلهٔ قبل را در یک اکتاو قرار داد؛ زیرا در فرآیند طبقه‌بندی، کارکردن با یک اکتاو بسیار ساده‌تر است. در موسیقی، اکتاو فاصلهٔ دو نت هم‌نام متوالی است. اگر بسامد دو نت متوالی که با هم یک اکتاو فاصله دارند، اندازه‌گیری شود، بسامد یکی دو برابر دیگری خواهد بود. به عبارت دیگر، نت‌هایی که با فاصله اکتاو از هم قرار می‌گیرند، تشکیل یک تصاعد هندسی را می‌دهند. برای این منظور فاصلهٔ فرکانسی بین ۲۲۰ تا ۴۴۰ هرتز انتخاب شده است. نت f_i به‌وسیلهٔ رابطه (۵) در این اکتاو قرار می‌گیرد:

$$FL(f_i) = \begin{cases} \dfrac{f_i}{2^{ceil\left(\log_2 \frac{f_i}{440}\right)}}, & f_i > 440 \\ f_i \times 2^{ceil\left(\log_2 \frac{220}{f_i}\right)}, & f_i < 220 \\ f_i, & \text{otherwise.} \end{cases} \qquad (۵)$$

[1] Discrete Fourier Transform (DFT)
[2] Fast Fourier Transform (FFT)
[3] Subtractive Clustering

پس از آن، کلیهٔ فرکانس‌های منتقل‌شده به یک اکتاو را نسبت به فرکانس اول بازه (۲۲۰ هرتز) به حوزه cent می‌بریم.

پس از آن، کلیهٔ فرکانس‌های منتقل‌شده به یک اکتاو را نسبت به فرکانس اول بازه (۲۲۰ هرتز) به حوزه cent می‌بریم. cent، یکی از واحدهای اندازه‌گیری در موسیقی است. در اعتدال مساوی دوازده نتی (روش کوک‌کردن سازها) هر اکتاو را به دوازده قسمت تقسیم می‌کنند که هر کدام از قسمت‌ها صد cent طول دارد. به‌طورمعمول از cent برای نشان‌دادن فواصل کوچک جهت مقایسهٔ اندازه فاصله‌ها در سامانه‌های متفاوت کوک‌بندی استفاده می‌شود. در واقع، فاصله یک cent، آن‌قدر کم است که نمی‌توان آن را میان نت‌های متوالی شنید. در حوزهٔ cent کلیهٔ فرکانس‌های بین دو بازه (برای مثال در اینجا ۲۲۰ تا۴۴۰ هرتز) محاسبه می‌شوند. تمام نت‌ها نسبت به فرکانس۲۲۰ هرتز از طریق رابطه (۶) به واحد cent تبدیل می‌شوند:

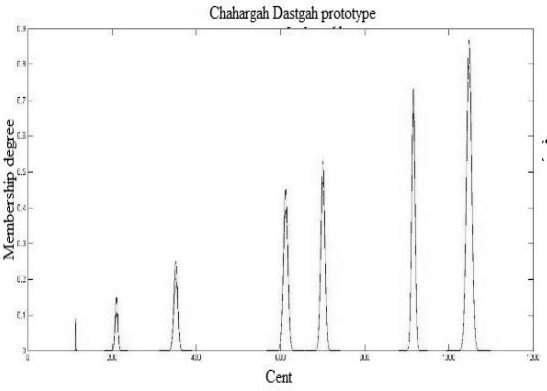
$$cent = 1200\log_2\left(\frac{FL}{220}\right). \qquad (۶)$$

پس از اینکه نت‌ها در یک اکتاو قرار گرفتند و تبدیل به حوزهٔ cent نیز صورت پذیرفت، فاصلهٔ ماهالانوبیس[1] اعمال می‌شود تا مشخص شود هر نقطه مربوط به کدام نت موسیقی است. فاصلهٔ بین فریم‌های مجاور از رابطهٔ زیر محاسبه می‌شود:

که در آن ماتریس M ماتریس کوواریانس است و از واریانس مقادیر موجود در فریم موردنظر محاسبه می‌شود. هرگاه در فریم p_i یک نت جدید آغاز شده باشد، فاصلهٔ بین دو فریم مجاور از یک حد آستانه بزرگ‌تر خواهد بود (در اینجا حد آستانه برابر با ۰/۲۲ می‌باشد). در مرحلهٔ بعد، عملیات فازی‌سازی صورت می‌گیرد. در این مرحله، باید اطلاعات سیگنال ناشناخته‌ی ورودی و همچنین، اطلاعات تئوری دستگاه‌های مرجع را به حوزهٔ منطق فازی ببریم. جهت تهیه اطلاعات تئوری، بهترین و کامل‌ترین قطعه را از هر دستگاه که دارای هفت نت اصلی موسیقی (هفت ساختار فازی) باشد، پس از پردازش‌های فازی‌سازی به‌عنوان الگوی مرجع در نظر می‌گیریم. با توجه به پنج دسته اصلی دستگاه موسیقی، پنج الگوی فازی به‌عنوان اطلاعات تئوری در نظر گرفته می‌شوند. درنهایت، برای هر کدام از اطلاعات تئوری و اطلاعات سیگنال ورودی یک سری دسته‌های فازی گوسین، به‌صورت شکل (۳) به‌دست می‌آید. از معیار شباهت فازی (FSM)[2] جهت محاسبهٔ میزان شباهت بین الگوی دستگاه‌ها (نمونه‌ی اصلی) و الگوی ناشناختهٔ ورودی استفاده می‌شود. برای به‌دست‌آوردن مقدار شباهت فازی، الگوریتم‌های زیادی ارائه شده است. در اینجا، از روش JSM[3] که توسط *Wu و همکارانش* تعریف شده [14]، استفاده شده است. این روش، سریع‌ترین الگوریتم بین FSMها است که به‌صورت رابطه (۸) تعریف می‌شود:

$$\tilde{S}(\tilde{A},\tilde{B}) = \frac{\int_X \min\left(\bar{\mu}_{\tilde{A}}(x),\bar{\mu}_{\tilde{B}}(x)\right)dx + \int_X \min\left(\underline{\mu}_{\tilde{A}}(x),\underline{\mu}_{\tilde{B}}(x)\right)dx}{\int_X \max\left(\bar{\mu}_{\tilde{A}}(x),\bar{\mu}_{\tilde{B}}(x)\right)dx + \int_X \max\left(\underline{\mu}_{\tilde{A}}(x),\underline{\mu}_{\tilde{B}}(x)\right)dx}, \qquad (۸)$$

که در آن، $\tilde{A}$ تابع عضویت ثانویه مجموعه اطلاعات سیگنال ورودی الگو، $\tilde{B}$ تابع عضویت ثانویه مجموعه اطلاعات سیگنال ناشناخته ورودی در دامنه x، $\bar{\mu}$ حد بالای تابع عضویت ثانویه فازی نوع دو، و $\underline{\mu}$ حد پایین تابع عضویت ثانویه فازی نوع دو است. x نیز دامنهٔ اطلاعات (در اینجا صفر تا ۱۲۰۰) است [2].



(شکل-۳): الگوی فازی دستگاه چهارگاه که از هفت مجموعهٔ فازی نوع دو تشکیل شده است.

(Figure-3): Fuzzy similarity prototype for Chahargah Dastgah which consists of 7 FSST2.

فاصله بین دو مجموعهٔ فازی به‌صورت زیر تعریف می‌شود:

$$\tilde{D}(\tilde{A},\tilde{B}) = 1 - \tilde{S}(\tilde{A},\tilde{B}). \qquad (۹)$$

[1] Mahalanobis Distance

[2] Fuzzy Similarity Measure (FSM)
[3] Jaccard Similarity Measure (JSM)

میانگین وزن‌دار فازی[1] (FWA) به‌وسیلهٔ رابطه (۱۰) محاسبه می‌شود:

$$y = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}. \qquad (۱۰)$$

درنهایت، جهت جداسازی نهایی، فرض می‌کنیم سیگنال ناشناخته ورودی تعدادی مؤلفه IT2FS دارد. از طریق فرمول (۸) تشابه بین IT2FSهای سیگنال ورودی و IT2FSهای تک‌تک دستگاه‌ها به‌دست می‌آید و در ماتریس تشابه قرار می‌گیرد. در مرحله بعد، مقدار بیشینهٔ هر سطر این ماتریس تشابه محاسبه می‌شود؛ سپس، این مقادیر بیشینه و وزن‌های نت‌های هر دستگاه در رابطه (۱۰) قرار داده شده و میانگین y برای هر دستگاه به‌دست می‌آید. هرگاه میانگین محاسبه‌شده مربوط به هر دستگاه بیشترین باشد، سیگنال ورودی ناشناخته مربوط به همان دستگاه است. رابطهٔ جداسازی به‌صورت زیر است:

$$L^* = \arg\max_l \left(FWA(Z_m^{-l}, W_m^l) \right). \qquad (۱۱)$$

۳- سامانه پیشنهادی جهت طبقه‌بندی دستگاه‌های موسیقی سنتی ایرانی مبتنی بر ویژگی‌های برگرفته از فرکانس گام

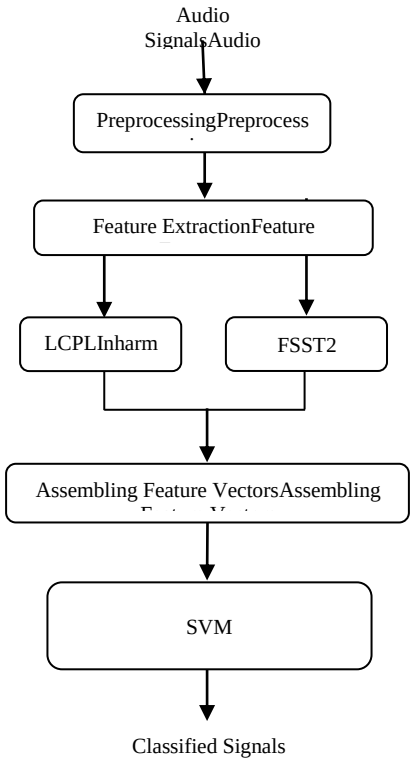
جهت طبقه‌بندی از الگوریتم‌های کلاسیک آموزش ماشینی، مانند شبکه‌های عصبی[2] (ANN)، شبکه‌های پرسپترون چندلایه، توابع RBF[3]، مدل‌های مارکوف پنهان[4] (HMM)، الگوریتم‌های KNN[5]، مدل‌های گوسی مختلط (GMM) و آنالیز تفکیک خطی (LDA)[6] استفاده می‌شود. در این پژوهش، ابتدا، روش نوینی مبتنی بر استخراج ویژگی‌های خاص برای دستگاه‌های موسیقی سنتی ایرانی، شامل دسته‌های شباهت فازی نوع۲[7] (FSST2) و ضرایب لاگرانژ لگاریتم فرکانس گام[8] (LCPL) پیشنهاد. سپس، از ماشین بردار پشتیبان چند کلاسی طبقه‌بندی استفاده می‌شود. نمودار جعبه‌ای این سامانه در شکل (۴) آمده است. همان‌طور که در شکل مشاهده می‌شود، ابتدا، روی سیگنال‌های ورودی پیش‌پردازش‌هایی، از جمله نمونه‌برداری، پنجره‌بندی و غیره، صورت می‌گیرد. سیگنال‌های پیش‌پردازش‌شده جهت استخراج ویژگی در ادامهٔ روند به‌کار برده می‌شوند. استخراج ویژگی، در مرحله بعد صورت می‌پذیرد. در سامانه پیشنهادی، از ترکیب دو ویژگی FSST2 و LCPL استفاده می‌شود. ویژگی FSST2 از سامانه سجاد عبدلی [2] برگرفته شده است. ویژگی دوم نیز با اعمال تغییرات و پردازش‌هایی که در ادامهٔ بحث بیان خواهد شد، بر روی فرکانس گام سیگنال به‌دست می‌آیند. پس از استخراج این دو ویژگی، بردار ویژگی از ترکیب دو ویژگی مطرح‌شده تشکیل شده و از آنها به‌منظور آموزش سامانه جداساز ماشین بردار پشتیبان استفاده می‌شود. در مرحله آزمون، سیگنال‌های موسیقی ناشناخته ورودی به سامانه جداساز ماشین بردار پشتیبان اعمال شده، و طبقه‌بندی صورت می‌گیرد.

۳-۱- ویژگی FSST2

ایده ویژگی دسته‌های شباهت فازی نوع۲ (FSST2) از سامانه سجاد عبدلی [2] گرفته شده که در بخش ۲-۲ به‌تفصیل توضیح داده شده است. ویژگی FSST2 در واقع با اعمال تغییراتی بر روی فرکانس گام موسیقی مورد نظر در حوزهٔ منطق فازی به‌دست می‌آید. نمودار جعبه‌ای مراحل استخراج این ویژگی در شکل (۵) آمده است. همان‌طور که در نمودار جعبه‌ای مشخص است، نخستین مرحله، استخراج فرکانس پایهٔ سیگنال ورودی است. برای این منظور از نرم‌افزار PRAAT [12] استفاده شده است. در مرحلهٔ دوم، با استفاده از روش خوشه‌بندی کاهشی [13] نت‌های اضافی حذف می‌شوند و طبقه‌بندی نت صورت می‌پذیرد. پس از آن تمام مراکز استخراج‌شده در مرحلهٔ قبل را در یک اکتاو قرار می‌دهیم؛ زیرا در فرآیند طبقه‌بندی، کارکردن با یک اکتاو بسیار ساده‌تر است. در اینجا، فاصلهٔ فرکانسی، بین ۲۲۰ هرتز تا ۴۴۰ هرتز انتخاب شده و از طریق رابطه (۵) کلیه فرکانس‌ها به این اکتاو منتقل می‌شوند. سپس، این فرکانس‌ها را نسبت به فرکانس نخست بازه (۲۲۰ هرتز) به حوزهٔ cent تبدیل می‌کنیم (رابطه ۶) و بعد از آن فاصله ماهالانوبیس را بین فریم‌های مجاور محاسبه می‌کنیم (رابطه ۷). در مرحلهٔ بعد، عملیات فازی‌سازی صورت می‌گیرد. در این مرحله، باید اطلاعات سیگنال ناشناختهٔ ورودی و همچنین، اطلاعات تئوری دستگاه‌های مرجع را به حوزهٔ منطق فازی ببریم. جهت تهیه اطلاعات تئوری، از هر دستگاه بهترین و کامل‌ترین قطعه که دارای هفت نت اصلی موسیقی (هفت

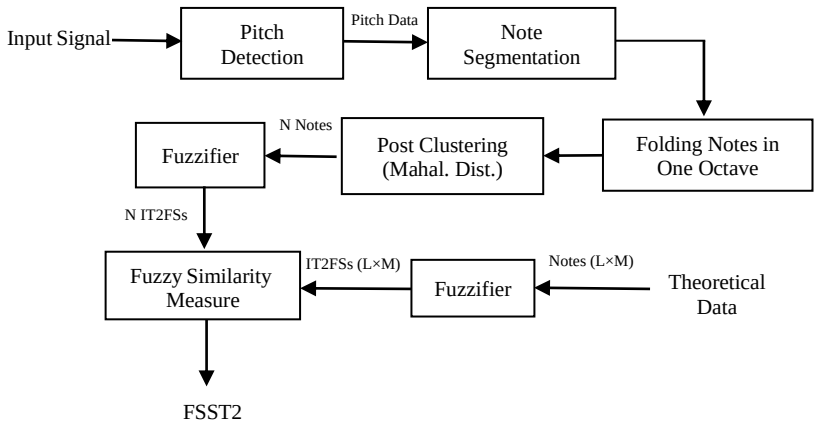
[1] Fuzzy Weighted Average (FWA)
[2] Artificial Neural Network (ANN)
[3] Radial Basis Function (RBF)
[4] Hidden Markov Model (HMM)
[5] K-Nearest Neighbors (KNN)
[6] Linear Discriminant Analysis (LDA)
[7] Fuzzy Similarity Sets Type 2 (FSST2)
[8] Lagrange coefficients of pitch logarithm (LCPL)

ساختار فازی) است، پس از پردازش‌های فازی‌سازی، به‌عنوان الگوی مرجع انتخاب می‌شود. با توجه به پنج دسته اصلی دستگاه موسیقی، پنج الگوی فازی به‌عنوان اطلاعات تئوری در نظر گرفته می‌شوند. در آخرین مرحله استخراج ویژگی برای هر دستگاه موسیقی، میزان شباهت فازی (FSM) سیگنال ورودی با این مرجع (رابطه ۸)، به‌عنوان ویژگی جدید در مقاله حاضر محاسبه می‌شود.



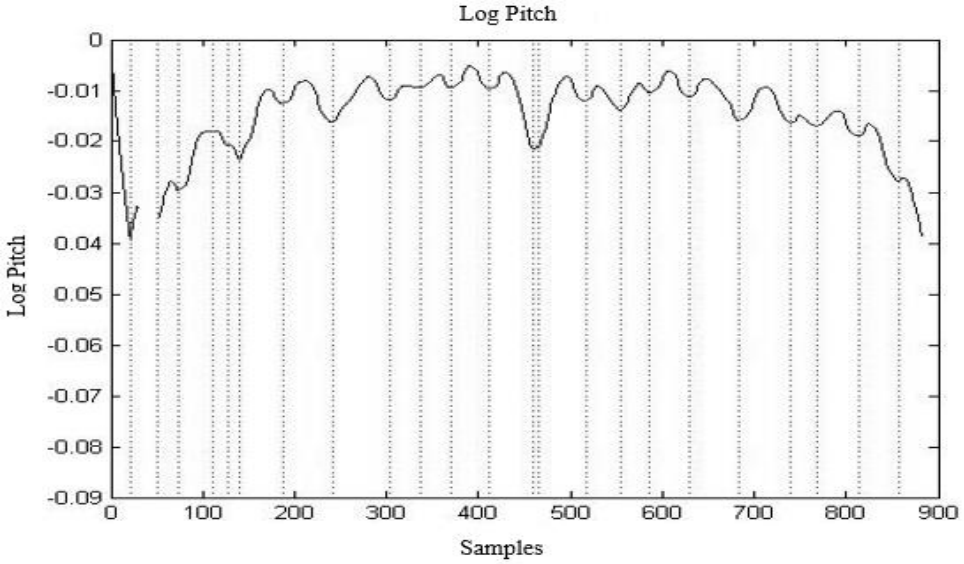
(شکل-۴): نمودار جعبه‌ای سامانهٔ پیشنهادی جهت دسته‌بندی دستگاه‌های موسیقی سنتی ایرانی مبتنی بر ترکیب دو ویژگی FSST2 و LCPL.

(Figure-4): The block diagram of the proposed system for the classification of Iranian traditional music Dastgahs based on the mixture of FSST2 and LCPL.



(شکل-۵): بلوک‌دیاگرام استخراج ویژگی FSST2 برگرفته از سیستم عبدلی [۲].

(Figure-5): The block diagram for extracting FSST2 feature based on the Abdoli's system [2].



(شکل-۶): نمونه‌ای از بخش‌بندی لگاریتم فرکانس گام برای یک فریم از قطعه‌ی موسیقی از دستگاه چهارگاه.

(Figure-6): An example of pitch logarythm segmentation for a frame of a piece of music from Chahargah Dastgah.

(جدول-۱): ماتریس ابهام سیستم عباسی لایق.

(Table-1): The confusion matrix of the Abbassi Layegh's system.

دستگاه‌ها	چهارگاه	همایون	ماهور و راست‌پنجگاه	سه‌گاه	شور و نوا
چهارگاه	2	0	8	0	0
همایون	1	4	3	2	0
ماهور و راست‌پنجگاه	4	2	4	0	0
سه‌گاه	0	3	1	5	1
شور و نوا	2	6	0	0	2

(جدول-۲): ماتریس ابهام سیستم عبدلی.

(Table-2): The confusion matrix of the Abdoli's system.

دستگاه‌ها	چهارگاه	همایون	ماهور و راست‌پنجگاه	سه‌گاه	شور و نوا
چهارگاه	5	2	1	1	1
همایون	3	5	1	1	0
ماهور و راست‌پنجگاه	1	4	4	0	1
سه گاه	1	3	1	5	0
شور و نوا	0	2	0	3	5

(جدول-۳): ماتریس ابهام سیستم پیشنهادی مبتنی بر ترکیب دو ویژگی FSST2 و LCPL.

(Table-3): The confusion matrix of the proposed system based on the combination of FSST2 and LCPL.

دستگاه‌ها	چهارگاه	همایون	ماهور و راست‌پنجگاه	سه‌گاه	شور و نوا
چهارگاه	7	2	0	1	0
همایون	1	8	1	0	0
ماهور و راست‌پنجگاه	0	2	6	0	2
سه گاه	1	2	0	7	0
شور و نوا	3	1	0	0	6

(جدول-۴): مقایسهٔ عملکرد سیستم‌های پایه و پیشنهادی در طبقه‌بندی دستگاه‌های موسیقی سنتی ایرانی بر حسب معیارهای Accuracy (٪)، Recall (٪)، Precision (٪)، F-measure (٪) و MCC ($[-1,+1]$).

(Table-4): The performance comparison of the baseline and proposed systems in classifying Iranian traditional music Dastgahs in terms of Accuracy (%), Recall (%), Precision (%), F-measure (%), and MCC ($[-1,+1]$).

سیستم طبقه‌بندی کننده		Accuracy	Recall	Precision	F-measure	MCC
عباسی لایق	چهارگاه	70	22.2	20	21	0.02
	همایون	66	26.6	40	32	0.1
	پنجگاه‌ماهور و راست	64	25	40	30.7	0.08
	گاه‌سه	86	71.4	50	58.8	0.5
	شور و نوا	82	66.6	20	30.7	0.29
	میانگین دستگاه‌ها	**73.6**	**42.3**	**34**	**34.6**	**0.21**
عبدلی	چهارگاه	80	50	50	50	0.37
	همایون	68	31.2	50	38.4	0.19
	پنجگاه‌ماهور و راست	82	57.1	40	47	0.37
	گاه‌سه	80	50	50	50	0.37
	شور و نوا	86	71.4	50	58.8	0.51
	میانگین دستگاه‌ها	**79.2**	**51.9**	**48**	**48.8**	**0.36**
پیشنهادی	چهارگاه	78	46.6	70	56	0.43
	همایون	96	100	80	88.8	0.87
	پنجگاه‌ماهور و راست	96	100	80	88.8	0.87
	گاه‌سه	98	100	90	94.7	0.93
	شور و نوا	88	70	70	70	0.62
	میانگین دستگاه‌ها	**91.2**	**83.3**	**78**	**79.7**	**0.74**

۲-۳- ویژگی LCPL

در این قسمت به توضیح نحوه استخراج ویژگی ضرایب لاگرانژ لگاریتم فرکانس گام (LCPL) می‌پردازیم. این ویژگی از سامانه Dehak [15] جهت تشخیص گوینده استخراج شده است. برای به‌دست‌آوردن این ویژگی، ابتدا، مقادیر فرکانس گام به‌وسیلهٔ برنامهٔ PRAAT استخراج می‌شوند در این برنامه، فرکانس گام به‌وسیلهٔ روش اتوکورولیشن محاسبه می‌شود، به‌نحوی‌که اطلاعات مربوط به قسمت‌های واکه (واکدار) جهت استفاده در ادامه روند استخراج این ویژگی حفظ شده، و اطلاعات به‌دست‌آمده برای قسمت‌های همخوان (بی‌واک) سیگنال حذف می‌شوند. سپس، مراحل زیر به‌ترتیب صورت می‌پذیرند:

الف- بخش‌بندی منحنی‌های لگاریتم فرکانس گام:

برای اینکه منحنی‌های فرکانس گام به‌عنوان یک واحد مدل شوند، تمامی منحنی‌های فرکانس گام پس از لگاریتم‌گیری به نواحی کوچک‌تر بخش‌بندی می‌شوند. این روش مبتنی بر آشکارکردن نقاط دره در این منحنی‌ها است. مختصات نقاط دره‌ها (به عبارتی، شماره فرکانس گام و مقدار لگاریتمی فرکانس) به‌عنوان مرزهای بخش‌بندی در نظر گرفته می‌شوند.

یک نمونه از این کار در شکل (۶) دیده می‌شود.

ب- تقریب منحنی‌های لگاریتم فرکانس گام:

منحنی بین هر دو دره به‌وسیلهٔ یک چندجمله‌ای لاگرانژ با درجه شش تقریب زده می‌شود. این منحنی‌ها به‌وسیلهٔ تابع $f(t)$ از رابطه (۹) تقریب زده می‌شود [15]:

که در آن، $p_i(t)$، i-امین جملهٔ لاگرانژ بوده و $M = 5$ انتخاب می‌شود. برای هر بخش شش ضریب برای منحنی لگاریتم فرکانس گام محاسبه شده و به‌عنوان یک بردار ویژگی شش‌بُعدی به‌دست می‌آید. این بردارهای ویژگی برای هر قطعه از سیگنال‌های موسیقی دستگاه‌ها استخراج شده و به جداساز ماشین بردار پشتیبان داده می‌شود.

۴- نتایج شبیه‌سازی

در بخش قبل، سامانه‌ای جهت طبقه‌بندی دستگاه‌های موسیقی سنتی ایرانی پیشنهاد شد. سپس، ویژگی‌های FSST2 و LCPL استخراج شده و از ماشین بردار پشتیبان چندرده‌ای برای طبقه‌بندی استفاده شد. یکی از مهم‌ترین قسمت‌هایی که در طبقه‌بندی باید به آن توجه کرد، ارزیابی سامانه طبقه‌بندی است. در این بخش، بر اساس معیارهای ارزیابی که در ادامه ارائه می‌شوند، میزان دقت و عملکرد الگوریتم‌های پایه و پیشنهاد شده مورد بررسی قرار می‌گیرند.

۴-۱- دادگان و شرایط آزمایشگاهی

در این پژوهش، جهت آزمون الگوریتم‌های پیشنهادی، از ۲۵۰ قطعه تک‌نوازی تار استاد علیزاده [16] استفاده شده است. برای هر دستگاه پنجاه قطعه در نظر گرفته شده که به‌طور تصادفی چهل قطعه در فرآیند آموزش سامانه جداساز، و ده قطعه نیز برای آزمون آن به‌کار گرفته می‌شود. از هر قطعه ده ثانیه انتخاب شده و با فرکانس نمونه‌برداری ۴۴۱۰۰ هرتز در محیط نرم‌فزار MATLAB خوانده می‌شود؛ سپس، فرآیند استخراج ویژگی صورت گرفته و عملیات آزمون انجام می‌شود. این مرحله با روش اعتبارسنجی تقاطعی (Cross Validation) پنج مرتبه صورت گرفته و درنهایت، از نتایج این پنج مرتبه میانگین گرفته می‌شود. برای سامانه جداساز، از جعبه‌ابزار LIBSVM [17] جهت رده‌بندی دستگاه‌ها استفاده شده است.

۴-۲- معیارهای ارزیابی

در این پژوهش، جهت ارزیابی عملکرد سامانه‌های پایه و پیشنهادی، از معیارهای Accuracy (برحسب درصد)، Recall (برحسب درصد)، Precision (برحسب درصد)، F-measure (برحسب درصد) [18] و MCC[1] [19] استفاده شده است. به‌عنوان یک ابزار مهم در ارزیابی الگوریتم‌های پایه و پیشنهادی، از ماتریس ابهام[2] نیز بهره برده می‌شود که با استفاده از آن، حالات مختلف طبقه‌بندی برای دستگاه‌ها به‌دست می‌آیند [18]. توصیف معیارهای Recall، Precision، Accuracy، F-measure و MCC در ادامه آمده است.

معیار Accuracy

مهم‌ترین معیار برای تعیین کارایی یک الگوریتم دسته‌بندی، معیار صحت (Accuracy) است. این معیار دقت کل یک دسته‌بند را محاسبه کرده و نشان‌دهنده میزان درستی دسته‌بندی داده‌ها نسبت به کل مجموعه‌داده‌ها است. این معیار از رابطهٔ زیر به‌دست می‌آید [18]:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}. \quad (۱۲)$$

که در آن:

TP: تعداد نمونه‌هایی که به‌درستی برچسب زده شده‌اند.

FP: تعداد نمونه‌هایی که متعلق به دستهٔ مورد نظر نیستند، اما به‌اشتباه در این دسته گنجانده شده‌اند.

TN: تعداد نمونه‌هایی که متعلق به دستهٔ مورد نظر نیستند و در این دسته هم قرار نگرفته‌اند.

FN: تعداد نمونه‌هایی که متعلق به دستهٔ مورد نظر هستند، اما به‌اشتباه در دستهٔ دیگری قرار گرفته‌اند.

معیار Recall

حساسیت (Recall) معیاری است که درصد نمونه‌های درست برچسب زده‌شده را نسب به کل نمونه‌هایی که باید در دستهٔ مورد نظر قرار گیرند، نشان می‌دهد [18]:

$$Recall = \frac{TP}{TP + FN}. \quad (۱۳)$$

معیار Precision

دقت (Precision) در مسائل دسته‌بندی، معیاری است که درصد نمونه‌های درست برچسب زده‌شده را نسب به کل نمونه‌هایی که در دسته مورد نظر قرار گرفته‌اند، نشان می‌دهد [18]. این معیار ارزیابی از رابطهٔ (۱۴) به‌دست می‌آید:

$$Precision = \frac{TP}{TP + FP}. \quad (۱۴)$$

معیار F-measure

زمانیکه Precision و Recall برای مسئلهٔ دسته‌بندی محاسبه شده‌اند، این دو معیار را می‌توان ترکیب کرده و به یک معیار واحد بنام F-measure رسید. رابطهٔ زیر نحوهٔ به‌دست‌آوردن این معیار را نشان می‌دهد [18]:

$$F\text{-}measure = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \quad (۱۵)$$

این معیار به‌عنوان متوسط هارمونیک[3] دو نسبت Precision و Recall تعبیر و تفصیر می‌شود.

[1] Mathews Correlation Coefficient (MCC)
[2] Confusion Matrix
[3] Harmonic Mean

معیار MCC

MCC به‌طوراصولی به‌عنوان یک معیار کیفی جهت بررسی نحوه رده‌بندی دوتایی است؛ اما برای دسته‌بندی با تعداد دسته بیش از دو نیز کاربرد دارد [۱۹]. معیار MCC درواقع، همبستگی دستجات واقعی را با برچسب‌های پیش‌گویی‌شده به‌وسیلهٔ دسته‌بند اندازه‌گیری می‌کند. MCC عددی بین ۱- و ۱ است، به‌نحوی‌که ۱+ بیان‌گر بالاترین میزان کارایی و ۱- بیان‌گر پایین‌ترین میزان کارایی است. این معیار بهترین راه برای تحلیل ماتریس ابهام است، زیرا در این معیار تمام پارامترهای مثبت و منفی TP، TN،FP و FN لحاظ شده است. معیار MCC به‌وسیلهٔ رابطه زیر محاسبه می‌شود:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \quad (۱۶)$$

۴-۳- مقایسه نتایج ارزیابی سیستم‌های پایه و پیشنهادی

نتایج شبیه‌سازی سامانه‌های مبنای عباسی لایق و سجاد عبدلی و همچنین، سامانه پیشنهادی در جدول‌های (۱)، (۲)، (۳) و (۴) آمده است. جداول (۱)، (۲) و (۳) ماتریس‌های ابهام نتایج آزمون سامانه‌های پایه و پیشنهادی را نشان می‌دهند. در این سامانه‌ها، از هر دستگاه ده قطعه جهت آزمون استفاده شده است. اعداد روی قطر اصلی ماتریس ابهام بیان‌گر این هستند که در هر دستگاه چه تعداد از نمونه‌ها درست تشخیص داده شده‌اند؛ در حالی‌که اعداد روی هر سطر خارج از قطر اصلی نشان می‌دهند چه تعداد از نمونه‌های آزمون متعلق به دستگاه مورد نظر، به‌اشتباه برچسب دستگاه دیگری را خورده‌اند. همان‌طور که در این جداول مشاهده می‌شود، بالاترین مقدار اعداد قطر اصلی، مربوط به سامانه پیشنهادی است که نشان‌دهندهٔ برتری عملکرد آن نسبت به سامانه‌های پایه است.

جدول (۴) مقادیر به‌دست‌آمده از معیارهای مختلف ارزیابی را در هر سه سامانه پایه و پیشنهادی نشان می‌دهد. علاوه‌بر آن میانگین کلی مقادیر ارزیابی بین تمام دستگاه‌ها نیز برای هر معیار ارزیابی در این جدول محاسبه شده است. همان‌طور که مشاهده می‌شود، در همه معیارها عملکرد سامانه پیشنهادی در طبقه‌بندی موسیقی سنتی ایرانی نسبت به هر دو سامانه پایه بیشتر است.

۵- نتیجه‌گیری

در این مقاله، سه سامانه جهت طبقه‌بندی دستگاه‌های موسیقی سنتی ایرانی، شامل دو سامانه مبنای عباسی لایق و سجاد عبدلی و سامانه پیشنهادی پیاده‌سازی شده‌اند. در سامانه پایه عباسی لایق، از ترکیب چهار ویژگی فرکانس گام، ضرایب ناموزونی، انحراف معیار استاندارد مرکز طیفی، و MFCC، و دسته‌بند ماشین بردار پشتیبان استفاده شده است.

سامانه پایه سجاد عبدلی از ویژگی فرکانس گام استخراج‌شده از داده ورودی به‌منظور طبقه‌بندی دستگاه‌ها به‌وسیلهٔ جداساز منطق فازی نوع دو استفاده کرده است. در سامانه پیشنهادی، با ترکیب دو ویژگی FSST2 و LCPL، که هر دو از فرکانس گام به‌دست آمده‌اند، عمل جداسازی دستگاه‌های موسیقی به‌وسیلهٔ ماشین بردار پشتیبان صورت می‌پذیرد. با مقایسهٔ نتایج دو سامانه پایه و سامانه پیشنهادی از نظر معیارهای مختلف ارزیابی، می‌توان به بالابودن میزان کارایی سامانه پیشنهادی در جداسازی نسبت به دو سامانه پایه پی برد. این موضوع نشان می‌دهد که ویژگی‌هایی که بر پایهٔ فرکانس گام استخراج می‌شوند، تأثیر بسیار بالاتری نسبت به ویژگی‌های مرسوم در جداسازی موسیقی ایرانی دارند.

۶- مراجع 6- References

[1] M. Abbasi Layegh, S. Haghipour, and Y. Najafi Sarem, "Classification of the Radif of Mirza Abdollah a canonic repertoire of Persian music using SVM method," *Gazi University Journal of Science Part A: Engineering and Innovation*, vol. 1, no. 4, pp. 57-66, 2013.

[2] S. Abdoli, "Iranian traditional music Dastgah classification," in *12th International Society for Music Information Retrieval Conference (ISMIR)*, Miami, Florida, USA, 2011, pp. 275-280.

[3] N. Darabi, N. Azimi, and H. Nojumi, "Recognition of Dastgah and Maqam for Persian music with detecting skeletal melodic models," in *the second annual IEEE BENELUX/DSP Valley Signal Processing Symposium*, 2006.
in *the second annual IEEE BENELUX/DSP Valley Signal Processing Symposium*, 2006.

[4] A. Ghosal, R. Chakraborty, B.C. Dhara, and S.KSaha, "Music classification based on MFCC variants and amplitude variation pattern: a hierarchical approach," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 5, no. 1, pp. 131-150, March 2012.

[5] J. Wülfing and M. Riedmiller, "Unsupervised learning of local features for music classification," in *13th International Society for Music Information Retrieval Conference (ISMIR)*, Portugal, 2012, pp.139-144.

[6] GTZAN Genre Collection, 2002. [online]. Available: http://marsyas.info/downloads/datasets.html [Accessed: July 28, 2021].

[7] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293-302, 2002.

[8] Y. Song, S. Dixon, and M.T. Pearce, "Evaluation of musical features for emotion classification," in *13th International Society for Music Information Retrieval (ISMIR)*, 2012, pp. 523-528.

[9] O. Lartillot, "MIRtoolbox 1.6.1," Aalborg University, Denmark, Department of Architecture, Design and Media Technology, December 7, 2014. [online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm [Accessed: July 28, 2021].

[10] V.N. Vapnik, "An overview of statistical learning theory," IEEE *Transactions on Neural Networks*, vol. 10, no. 5, pp.988-999, 1999.

[11] C.W. Hsu and C.J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415-425, 2002.

[12] P. Boersma and D. Weenink, Institute of Phonetics Sciences, University of Amsterdam, Version 6.1.51, 2021. [Online]. Available: http://www.praat.org/ [Accessed: July 28, 2021].

[13] S.L. Chiu, "Fuzzy model identification based on cluster estimation," *Journal of Intelligent & Fuzzy Systems*, vol. 2, no. 3, pp. 267-278, 1994.

[14] D. Wu and J.M. Mendel, "A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets," *Information Sciences*, vol. 179, no. 8, pp. 1169-1192, 2009.

[15] N. Dehak, P. Dumouchel, and P. Kenny, "Modeling prosodic features with joint factor analysis for speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 7, pp. 2095-2103, 2007.

[16] H. Alizadeh. [Online]. Available: http://www.hosseinalizadeh.net [Accessed: July 28, 2021].

[17] C.C. Chang and C.J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1-27, 2011.

[18] D.M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *arXiv preprint arXiv:2010. 16061*, 2020.

[19] G. Jurman, S. Riccadonna, and C. Furlanello, "A comparison of MCC and CEN error measures in multi-class prediction," *PLOS ONE*, vol. 7, no. 8, 2012. [Online]. Available: https://doi.org/10.1371/journal.pone.0041882. [Accessed: July 28, 2021].

[20] K. Choi, G. Fazekas, M. Sandler, and K. Cho, "Convolutional recurrent neural networks for music classification," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 2392-2396.

[21] Y.M. Costa, L.S. Oliveira, and C.N. Silla Jr, "An evaluation of convolutional neural networks for music classification using spectrograms," *Applied Soft Computing*, vol. 52, pp. 28-38, 2017.

[22] J. Lee and J. Nam, "Multi-level and multi-scale feature aggregation using sample-level deep convolutional neural networks for music classification," *arXiv preprint arXiv:1706.06810*, 2017.

[23] B. Beigzadeh and M. Belali Koochesfahani, "Classification of Iranian traditional musical modes (DASTGÄH) with artificial neural network," *Journal of Theoretical and Applied Vibration and Acoustics*, vol. 2, no. 2, pp. 107-118, 2016.

[24] P. Heydarian, "Automatic recognition of Persian musical modes in audio musical signals," Ph.D. dissertation, London Metropolitan University, 2016.

[۲۵] حبیبی اقدم حامد، همایون پور محمد مهدی، " شناسایی خودکار سبک موسیقی"، مجله پردازش علائم و داده ها، دوره ۷، شماره ۱، صفحات ۳۳-۵۲، ۱۳۸۹.

[25] H. Habibi Aghdam, M.M. Homayoun pour, "Automatic Recognition of Music Genre," *Journal of Signal and Data Processing*, vol. 7, Issue 1, pp. 33-52, 2010.

مسعود گراوانچی‌زاده مدرک کارشناسی را در رشته مهندسی الکترونیک در سال ۱۳۶۵ از دانشگاه تبریز دریافت کرده است. وی سپس، مدارک کارشناسی ارشد و دکترا در رشته پردازش سیگنال، به‌ترتیب، در سال‌های ۱۳۷۴ و ۱۳۸۰ از دانشگاه بوخوم آلمان کسب کرده است. ایشان هم‌اکنون عضو هیأت علمی دانشکده مهندسی برق و کامپیوتر دانشگاه تبریز با مرتبه علمی دانشیاری هستند. زمینه‌های پژوهشی مورد علاقه ایشان بهبود کیفیت گفتار، مکان‌یابی و جداسازی سیگنال گفتار، پردازش سیگنال‌های تصادفی و پردازش سیگنال‌های دو گوشی است.

نشانی رایانامه ایشان عبارت است از:

geravanchizadeh@tabrizu.ac.ir

پریسا مبشری مدارک کارشناسی برق-الکترونیک و کارشناسی ارشد برق-مخابرات،گرایش سیستم، را در سال‌های ۱۳۸۸ و۱۳۹۵، به‌ترتیب، از دانشگاه‌های آزاد اسلامی واحد نجف آباد و علوم و تحقیقات آذربایجان شرقی دریافت کرده است. زمینه‌های پژوهشی مورد علاقه ایشان پردازش سیگنال‌های موسیقی، جداسازی سیگنال‌های گفتار و موسیقی و شنوایی ماشین هستند.

نشانی رایانامه ایشان عبارت است از:

p.m122@yahoo.com

هادی جمشیدی اوانکی مدرک کارشناسی خود را در ر شته مهندسی پزشکی، گرایش بیوالکتریک، در سال۱۳۹۲ از دانشگاه صنعتی سهند دریافت کرده است. وی سپس، مدرک کارشناسی ارشد خود را در رشته مهندسی مخابرات، گرایش سیستم، در سال ۱۳۹۵ از دانشگاه تبریز دریافت کرده است. زمینه‌های پژوهشی مورد علاقهٔ ایشان بهبود کیفیت و قابلیت فهم گفتار، علوم شناختی و الگوریتم‌های هوش محاسباتی با کاربرد در پردازش سیگنال است.

نشانی رایانامه ایشان عبارت است از:

Cfm.hadi@yahoo.com