پردیس علوم
دانشکده ریاضی، آمار و علوم کامپیوتر

تشخیص نوع و تعداد ساز در حوزه موسیقی سنتی ایرانی

نگارنده

سحرسادات شیرمردی

استاد راهنما: دکتر باقر باباعلی

پایان‌نامه برای دریافت درجه کارشناسی
در رشته علوم کامپیوتر

بهمن ۱۴۰۰

چکیده

شناسایی و دسته‌بندی خودکار دستگاه‌ها، سازها و گوشه‌های موسیقی سنتی ایرانی یکی از شاخه‌های حوزه بازیابی اطلاعات موسیقی محسوب می‌شود که مورد توجه چندی از پژوهشگران قرار گرفته است؛ اما به متاسفانه اندازه کافی به آن پرداخته نشده است. این در حالی است که در دنیا پژوهش‌های بسیاری در حوزه پردازش سیگنال موسیقی به منظور بازیابی محتوای موسیقی انجام شده است و این حوزه مورد توجه بسیاری از محققین قرار گرفته است. اما پژوهش‌های انجام شده در زمینه پردازش موسیقی سنتی ایرانی بسیار اندک است که یکی از دلایل آن می‌تواند در دسترس نبودن دادگان جامع و متنوع باشد. از آن جایی که اکثر دادگان‌های تشخیص آلات موسیقی بر روی آلات موسیقی غربی تمرکز دارند، مطالعه و ارزیابی حوزه تشخیص آلات موسیقی سنتی ایرانی برای محققان دشوار است. در عمده پژوهش‌هایی که تاکنون منتشر شده اند، صرفاً بر اساس فواصل بین نغمه‌های صوتیِ پنج دستگاه اصلی، سعی در تفکیک و شناسایی خودکارِ این دستگاه‌ها از یکدیگر داشته‌اند. از آنجا که دسته‌بندی دستگاه‌ها از اصالت لازم برخوردار نبوده و در این خصوص بینِ نظریه پردازان و موسیقی‌دان‌ها، از نظر تعداد دستگاه و مرز بین آن‌ها اتفاق نظر وجود ندارد، هنوز انجام پژوهش بیشتر و با روش‌های نوین در این زمینه نیاز است.

همانطور که بیان شد، یکی از دلایل اساسی کمبود پژوهش‌های زمینه موسیقی سنتی ایران را می‌توان نبود دادگان برشمرد. همچنین مشکل دیگری که بر سر راه این پژوهش‌ها هست، مسئله برچسب‌گذاری و حاشیه‌نویسی دستی است که بسیار پرهزینه، زمانبر و نیازمند نیروی متخصص در حوزه موسیقی است. لذا این کار پژوهشی، با هدف گردآوری یک دادگان جامع و متنوع برای یک مسئله اساسی حوزه موسیقی سنتی ایرانی، یعنی تشخیص ساز و در ادامه پایه گذاری یک بستر پایه برای این منظور، انجام شده است.

کلمات کلیدی: سازهای سنتی ایرانی، دادگان‌های موسیقی سنتی ایرانی، موسیقی سنتی ایرانی، یادگیری ماشین، یادگیری خودنظارتی، یادگیری متضاد، ارزیابی تشخیص آلات و دستگاه‌های موسیقی،

سپاسگزاری

از جناب آقای دکتر باقر باباعلی به‌عنوان استاد راهنما که همواره بنده را مورد لطف خود قرار داده‌اند، کمال تشکر را دارم؛ چرا که بدون راهنمایی‌های ایشان تامین این پایان‌نامه بسیار مشکل می‌نمود. همچنین با سپاس از شرکت نواک و بیپ‌تونز، به دلیل یاری‌ها و همکاری‌های بی‌چشم‌داشت ایشان که بسیاری از سختی‌ها را برایم آسانتر نمودند تا این پایان‌نامه را به پایان برسانم.

پیشگفتار

دسته‌بندی خودکار و کارآمد موسیقی از اهمیت حیاتی برخوردار است و مبنایی برای کاربردهای پیشرفته مختلف هوش مصنوعی در حوزه موسیقی است. تشخیص ساز موسیقی، وظیفه شناسایی نوع ساز به کار رفته به واسطه صدای آن است. این امر، گاهی بسیار ساده و گاهی بسیار پیچیده و دشوار است. به طور مثال تشخیص صدای تار و سه تار از هم، و یا تفکیک انواع سازهای کوبه‌ای از روی صدا گاهی حتی برای موسیقی‌دان‌ها هم بسیار دشوار است و با تغییر پارامترهایی مانند کوک سازها، سبک نواختن و ساختار فیزیکی سازها و غیره ممکن است دشوارتر هم بشود. این صدا، که ارتعاشات صدا نیز نامیده می‌شود، توسط مدل برای مطابقت با کلاس‌های ساز اعمال می‌شود.

برای پردازش سیگنال موسیقی به منظور بازیابی محتوای موسیقی و بدون ساده‌سازی مسئله و محدود کردن دادگان به قطعات تکنوازی و به روش یادگیری با نظارت، به دادگانی با حجم بالا، یکدست از نظر تنوع سازهای مختلف و با برچسب‌گذاری دقیق، نیاز است که تهیه آن بسیار زمان‌بر، پرهزینه و دشوار است. مزید بر این علت، فرآیند برچسب‌گذاری قطعات هم باید توسط متخصص این زمینه صورت گیرد که خود باعث دشوارتر شدن تهیه دادگان می‌شود و همچنان هم درصدی خطا وجود خواهد داشت. بنابراین، متاسفانه پژوهش‌های منتشرشده در حوزه پردازش رایانه‌ای موسیقی سنتی ایرانی بسیار ناچیز است. این پژوهش‌ها، که شامل شناسایی و دسته‌بندی خودکار دستگاه‌ها، سازها و گوشه‌های موسیقی سنتی ایرانی هستند، بیش از یک دهه است توجه پژوهشگران را به خود جلب کرده‌اند و یکی از شاخه‌های حوزه بازیابی اطلاعات موسیقی محسوب می‌شوند.

به طور کلی، اصلی‌ترین محدودیت‌ها و دلایل ناکارآمدی پژوهش‌های این حوزه را می‌توان ناشی از نبود پایگاه داده منسجم برای موسیقی ایرانی، انجام پژوهش‌ها به صورت موازی و مجزا و نداشتنِ دانش کافی پژوهشگران از مبانی نظری موسیقی سنتی ایرانی دانست. هدف از این پروژه، بررسی عمیق‌تر مبانی نظری موسیقی سنتی ایرانی، جمع آوری دادگان جامع و ساختن بستری مناسب، جهت انجام بهتر پژوهش در این حوزه بوده است[۲۵].

در این راستا، در ادامه به بررسی محدودیت‌های یادگیری با نظارت پرداخته شده است. خواهیم دید که با آن که یادگیری عمیق پیشرفت‌های بزرگی را در بسیاری از زمینه‌های موسیقی ممکن کرده است، به دلیل این که متکی بر دادگان‌های موسیقی برچسب‌دار است، معایب و محدودیت‌هایی دارد؛ زیرا حجم عظیمی از داده‌های بدون برچسب در هر ساعت تولید می‌شود که استفاده از این روش برای آن‌ها بهینه به نظر نمی‌رسد. علاوه بر این، مدل‌هایی که با دادگان‌های برچسب‌گذاری

شده آموزش داده می‌شوند، اغلب تحت تاثیر بایاس آن دادگان خاص قرار می‌گیرد. همچنین در کل، هوش و یادگیری، در مورد نگاشت ورودی به برچسب‌ها نیست و نیاز به راه حلی برای این مشکل، به شدت محسوس است. بدین منظور، یادگیری خودنظارتی، به عنوان نوعی یادگیری بی نظارت، به عنوان رویکردی جایگزین موفق عمل می‌کند.

در این پروژه از SimCLR [۷] استفاده شده و زنجیره بزرگی از داده‌های صوتی را برای ایجاد یک چارچوب ساده برای یادگیری متضاد خودنظارتی بازنمایش‌های موسیقایی به نام CLMR [۲۲] به کار رفته است که یک رویکرد اجرایی بر روی داده‌های موسیقی خام با نمایش دامنه زمانی است و برای یادگیری بازنمایش‌های مفید نیازی به برچسب ندارد. نمایش‌های CLMR با استفاده از مجموعه داده‌های خارج از دامنه قابل انتقال هستند، که نشان می‌دهد این روش، تعمیم‌پذیری قوی در دسته‌بندی موسیقی دارد. در نهایت، روش پیشنهادی یادگیری کارآمد داده را در دادگان‌های برچسب‌گذاری‌شده کوچک‌تر کاهش می‌دهد و بدین منظور از دادگان نوا[۱] که شامل تعداد زیادی قطعات تکنوازی از پنج ساز موسیقی سنتی ایرانی و برچسب‌دار هست، برای آموزش و تعمیم به دادگان جدید استفاده شده است.

فهرست مطالب

فصل ۱

مفاهیم مقدماتی

امروزه به طور معمول، وقتی بحث هوش مصنوعی مطرح می‌شود، منظور در اصل یادگیری ماشین است و آنچه در واقع انجام می‌شود یادگیری عمیق است و در یادگیری عمیق آنچه به طور معمول استفاده می‌شود یادگیری با نظارت است. در این بخش، هر یک از این مفاهیم به طور مختصر توضیح داده می‌شود.

۱.۱ یادگیری ماشین

هوش مصنوعی علم گسترده تقلید از توانایی‌های انسان است، و یادگیری ماشین زیرمجموعه خاصی از هوش مصنوعی است که به ماشین نحوه یادگیری را آموزش می‌دهد. در واقع، روشی برای تجزیه و تحلیل داده است که ساخت مدل تحلیلی را خودکار می‌کند. این شاخه از هوش مصنوعی، مبتنی بر این ایده است که سیستم‌ها می‌توانند از داده‌ها یاد بگیرند، الگوها را شناسایی کنند و با کمترین دخالت انسان تصمیم بگیرند؛ یعنی بدون نیاز به دستورات صریح از سمت انسان آموزش ببینند. با استفاده از شبیه‌سازی‌های مغز انسان، امید بر آن است که:

- استفاده از الگوریتم‌های یادگیری بسیار بهتر و آسان‌تر شود.
- پیشرفت‌های انقلابی در یادگیری ماشین و هوش مصنوعی رخ دهد.

دو روش از رایج‌ترین روش‌های یادگیری ماشین، یادگیری با نظارت و یادگیری بی نظارت هستند که در ادامه به طور مختصر به توضیح آن‌ها پرداخته شده است.

۱.۱.۱ یادگیری عمیق

یادگیری عمیق زیرمجموعه‌ای از یادگیری ماشین است که در اصل یک شبکه عصبی با سه لایه یا بیشتر است. این شبکه‌های عصبی تلاش می‌کنند تا رفتار مغز انسان را شبیه‌سازی کنند و به آن اجازه می‌دهند از مقادیر زیادی داده، آموزش ببینند. با آن که یک شبکه عصبی با یک لایه هم می‌تواند پیش‌بینی‌های تقریبی انجام دهد، لایه‌های پنهان اضافی می‌توانند به بهینه سازی و افزایش دقت کمک کنند. فناوری یادگیری عمیق در محصولات و خدمات روزمره (مانند دستیارهای دیجیتال، کنترل‌های تلویزیون با قابلیت صوتی و تشخیص تقلب در کارت اعتباری) و همچنین فناوری‌های نوظهور (مانند خودروهای خودران) به کار رفته است.

۲.۱ چندی از روش‌های یادگیری ماشین

۱.۲.۱ یادگیری بی نظارت

از ابتدا، دو نوع کار اساساً متفاوت در یادگیری ماشین وجود داشته است؛ اولین مورد یادگیری بی نظارت است. یادگیری بی نظارت برای داده‌هایی استفاده می‌شود که هیچ برچسبی ندارند. به سیستم پاسخ صحیح گفته نشده است و الگوریتم باید بفهمد که چه چیزی نشان داده شده است. هدف، کاوش داده‌ها و یافتن ساختاری در درون آن‌هاست.

فرض کنید $X = (x_1, ..., x_n)$ مجموعه‌ای از n مثال‌ها (یا نقطه‌ها) باشد، که در آن به ازای هر $i \in [n] := \{1, ..., n\}$ داشته باشیم $x_i \in \chi$. معمولاً فرض بر این است که نقاط به صورت مستقل با توزیع یکسان از یک توزیع مشترک در χ ترسیم می‌شوند. اغلب راحت است که ماتریسی$(n \times d)$ بعدی را که نقاط داده به عنوان ردیف‌های آن است، به صورت ذیل تعریف کنیم:

$$X = (x_i{}^{\tau})_{i \in [n]}^{\tau}$$

هدف یادگیری بی نظارت یافتن ساختار جالب در داده‌های X است. استدلال شده است که یادگیری بی نظارت اساساً مربوط به تخمین چگالی است که احتمالاً دلیل ایجاد X است[۶].

مدل ابتدا به خوشه‌بندی اولیه داده‌ها می‌پردازد، سپس با بررسی تراکم خوشه‌ها آن‌ها را تغییر می‌دهد و این افزایش و کاهش اندازه خوشه‌ها را تا جایی ادامه می‌دهد که به بهترین خوشه‌بندی ممکن دست یابد. یادگیری بی نظارت معمولاً الگوریتم‌های پیچیده‌تری را برای خوشه‌بندی به کار می‌گیرد. صورت‌های ضعیف‌تری نیز از یادگیری بی نظارت نیز وجود دارد؛ مانند تخمین کمیت، خوشه‌بندی، تشخیص داده پرت و کاهش ابعاد و غیره. تکنیک‌های رایج شامل نگاشت‌های خودسازمان‌دهی، خوشه‌بندی تاK نزدیک‌ترین همسایه و خوشه‌بندیK میانگین است. این الگوریتم‌ها برای تقسیم‌بندی موضوعات متنی، توصیه‌گرها و شناسایی داده‌های پرت استفاده می‌شوند.

۲.۲.۱ یادگیری با نظارت

وظیفه دوم یادگیری با نظارت است. الگوریتم‌های یادگیری با نظارت، با استفاده از داده‌های برچسب‌گذاری شده آموزش داده می‌شوند. داده‌های آموزش با برچسب‌هایی که نشان دهنده دسته‌ها هستند، برچسب‌گذاری شده‌اند. الگوریتم مجموعه‌ای از ورودی‌ها را به همراه خروجی‌های صحیح مربوطه دریافت می‌کند و با مقایسه خروجی بدست آمده با خروجی‌های صحیح و یافتن خطاها، آموزش می‌بیند و سپس مدل را مطابق با آن اصلاح می‌کند. یادگیری با نظارت از طریق روش‌هایی مانند دسته‌بندی، رگرسیون، پیش‌بینی و تقویت گرادیان، از الگوهایی برای پیش‌بینی مقادیر برچسب روی داده‌های بدون برچسب استفاده می‌کند.

یادگیری با نظارت معمولاً در برنامه‌هایی استفاده می‌شود که داده‌های گذشته رویدادهای احتمالی آینده را پیش‌بینی می‌کنند. تحقیقات انجام شده بر بروی موسیقی سنتی ایرانی تا کنون با به کارگیری این نوع یادگیری بوده است.

هدف یادگیری نگاشت از x به y، با توجه به مجموعه آموزشی ساخته شده از جفت $(x_i، y_i)$ است. در اینجا، $y_i \in Y$ برچسب‌ها یا اهداف مثال‌های x_i نامیده می‌شوند. اگر برچسب‌ها اعداد باشند، $y = (y_i)_{i\in[n]}^{\tau}$ بردار ستون برچسب‌ها را نشان می‌دهد. مجدداً، یک الزام استاندارد این است که جفت‌ها، $(x_i، y_i)$ ، به صورت مستقل با توزیع یکسان از توزیعی که در اینجا با برد $Y\times\chi$ است، نمونه‌برداری شوند. این رویکرد، خوش‌تعریف است؛ زیرا از طریق عملکرد پیش‌بینی آن در نمونه‌های آزمایشی، امکان ارزیابی یک نگاشت وجود دارد. هنگامی که $Y = \mathbb{R}$ یا $Y = \mathbb{R}^d$ (یا به طور کلی، زمانی که برچسب‌ها پیوسته باشند)، این روش رگرسیون نامیده می‌شود[۶].

دو خانواده از الگوریتم‌ها برای یادگیری با نظارت وجود دارد؛ روش‌های مولد و روش‌های تمایزی

روش‌های مولد

الگوریتم‌های مولد سعی می‌کنند چگالی کلاس شرطی $p(x|y)$ را با روش یادگیری بی نظارت مدل‌سازی کنند. سپس با اعمال قضیه بیز می‌توان چگالی پیش‌بینی را استنباط کرد:

$$p(y|x) = \frac{p(x|y)p(y)}{\int_y p(x|y)p(y)dy}$$

در واقع $p(x|y)p(y) = p(x, y)$ چگالی مشترک داده‌ها است که می‌توان از آن جفت‌های $(x_i، y_i)$ تولید کرد.

روش‌های تمایزی

الگوریتم‌های تمایزی تلاش نمی‌کنند تا چگونگی تولید x_i را تخمین بزنند، بلکه بر روی تخمین $p(y|x)$ تمرکز می‌کنند. برخی از روش‌های تمایزی حتی خود را به مدل‌سازی این که آیا $p(y|x)$

بزرگ‌تر یا کمتر از ۰.۵ است، محدود می‌کنند. نمونه‌ای از این روش، ماشین بردار پشتیبانی (SVM) است. استدلال شده است که مدل‌های تمایزی، مستقیم‌تر با هدف یادگیری با نظارت همسو هستند و بنابراین در عمل کارآمدتر هستند.

۳.۲.۱ یادگیری خودنظارتی

این روش به اندازه دو روش قبل، رایج و قدیمی نیست. یادگیری خودنظارتی به عبارتی، نوعی یادگیری بی نظارت است؛ زیرا از معیارهای مشابهی پیروی می‌کند و با داده‌هایی که هیچ برچسبی ندارند، سر و کار دارد. با این حال، به جای یافتن الگوهای سطح بالا و کلی برای خوشه‌بندی، یادگیری خودنظارتی تلاش می‌کند تا کارهایی را که به طور معمول توسط یادگیری با نظارت هدف قرار می‌گیرند، بدون هیچ گونه برچسب‌گذاری حل کند. چند مورد از از مزایای یادگیری خودنظارتی، عدم نیاز به مداخله انسان و مقیاس‌پذیر بودن و تعمین‌پذیری آن است.

یک مسئله مهم و شایان ذکر در یادگیری خودنظارتی این است که در آن از طریق تابع زیان جایگزین آموزش انجام می‌شود. در این روش، به منظور حفظ اطلاعات مفید، تابع زیان جایگزین با توجه به یک مسئله ابتدایی ساده صورت می‌گیرد. هدف از این رویکرد و ایده اصلی این است که بازنمایشی که برای مسئله خوب باشد، برای وظایف پایین‌دستی نیز مفید است. در واقع با پرداختن به وظایف پایین‌دستی و مسائل ساده، ماشین به یادگیری می‌پردازد.

۴.۲.۱ یادگیری نیمه‌نظارتی

یادگیری نیمه‌نظارتی، روشی بین یادگیری با نظارت و بی نظارت است. علاوه بر داده‌های بدون برچسب، الگوریتم از برخی از اطلاعات نظارتی استفاده می‌کند؛ اما نه لزوما برای همه نمونه‌ها. در این روش از حجم کمی از داده‌های برچسب‌گذاری شده به منظور یادگیری اولیه استفاده می‌شود. سپس از داده‌هاس آموزشی بدون برچسب یا با برچسب ساختگی برای آموزش استفاده می‌شود و در نهایت نتایج ترکیب شده و مدل بدست می‌آید. در این مورد، مجموعه داده $X = (x_i)_{i\in[n]}$ را می‌توان به دو بخش تقسیم کرد: نقاط $X_l := (x_1, ..., x_l)$ که برچسب های $Y_l := (y_1, ..., y_l)$ برای آن‌ها ارائه شده است و نقاط $X_u := (x_{l+1}, ..., x_{l+u})$ که برچسب های آن‌ها مشخص نیست[۶].

این نوع یادگیری را می‌توان ترکیبی از یادگیری با نظارت و بی نظارت دانست. یکی از رویکردهای رایج در این نوع یادگیری، ادغام الگوریتم‌های دسته‌بندی و خوشه‌بندی است.

۵.۲.۱ یادگیری متضاد

یادگیری متضاد یک رویکرد در یادگیری ماشین است که برای یادگیری ویژگی‌های کلی دادگان بدون برچسب، با آموزش مدل برای نقاط داده مشابه یا متفاوت استفاده می‌شود. یادگیری متضاد، بررسی

می‌کند که کدام جفت نقاط داده مشابه و متفاوت هستند تا ویژگی‌های سطح بالاتر در مورد داده‌ها را یاد بگیرد (حتی قبل از انجام وظایفی مانند دسته‌بندی). با یادگیری متضاد، می‌توان عملکرد مدل را به طور قابل توجهی بهبود بخشید؛ حتی زمانی که فقط کسری از مجموعه داده برچسب‌گذاری شده باشد.

۶.۲.۱ یادگیری بازنمایی

یادگیری بازنمایی، کلاسی از رویکردهای یادگیری ماشینی است که به سیستم اجازه می‌دهد تا بازنمایی‌های مورد نیاز را، برای تشخیص یا دسته‌بندی ویژگی‌ها، از داده‌های خام کشف و استخراج نماید.

در یادگیری بازنمایی، داده‌ها به ماشین فرستاده می‌شوند و ماشین به تنهایی بازنمایش را یاد می‌گیرد. این یادگیری، روشی برای تعیین نمایش داده‌ای از ویژگی‌ها، تابع فاصله و تابع شباهتی است که نحوه عملکرد مدل پیش‌بینی کننده را تعیین می‌کند. یادگیری بازنمایی با کاهش داده‌های با ابعاد بالا به داده‌های کم‌بعد کار می‌کند و کشف الگوها و ناهنجاری‌ها را آسان‌تر می‌کند و در عین حال درک بهتری از رفتار کلی داده‌ها ارائه می‌دهد.

اساساً، وظایف یادگیری ماشینی مانند دسته‌بندی، اغلب ورودی‌هایی را می‌طلبند که از نظر ریاضی و محاسباتی برای پردازش راحت باشد، که باعث ایجاد انگیزه در یادگیری بازنمایی می‌شود. داده‌های دنیای واقعی، مانند عکس‌ها، ویدیوها و غیره، در برابر تلاش‌ها برای تعریف کیفیت‌های خاص به‌صورت الگوریتمی مقاومت کرده‌اند. یک رویکرد این است که، به جای وابستگی به تکنیک‌های صریح، داده‌ها را برای چنین ویژگی‌ها یا بازنمایی‌هایی بررسی کنیم.

داده افزایی

داده افزایی در تجزیه و تحلیل داده‌ها روشی است که برای افزایش حجم داده‌ها با افزودن کپی‌های کمی متفاوت از داده‌های موجود یا داده‌های مصنوعی جدید تولید شده، استفاده می‌شود که به عنوان یک منظم کننده عمل می‌کند و به کاهش برازش بیش از حد در هنگام آموزش یک مدل یادگیری ماشینی کمک می‌کند. داده‌افزایی پیش از ورود داده به مدل انجام می‌شود. برای انجام این کار، دو رویکرد وجود دارد؛

- برون‌خط

در این رویکرد از قبل تمام تبدیلات لازم انجام می‌شود. در واقع اندازه دادگان اساساً افزایش پیدا می‌کند. این روش برای دادگان کم‌حجم مناسب است؛ به این دلیل که در این روش می‌توان افزایش انداره دادگان را با ضریبی برابر با تعداد تبدیلات انجام شده، به پایان رساند.

- برخط

در این رویکرد که به آن داده‌افزایی حین اجرا نیز گفته می‌شود، تبدیلات روی دسته کم حجمی از داده‌ها انجام شده و سپس به مدل یادگیری داده می‌شود. این روش برای دادگان با حجم زیاد ترجیح داده می‌شود. دلیل آن هم این است که داده‌افزایی روی تمام دادگان قابل اجرا نیست و به جای آن تبدیلات روی دسته‌های کوچکی که به مدل داده خواهند شد، انجام می‌شود.

فصل ۲

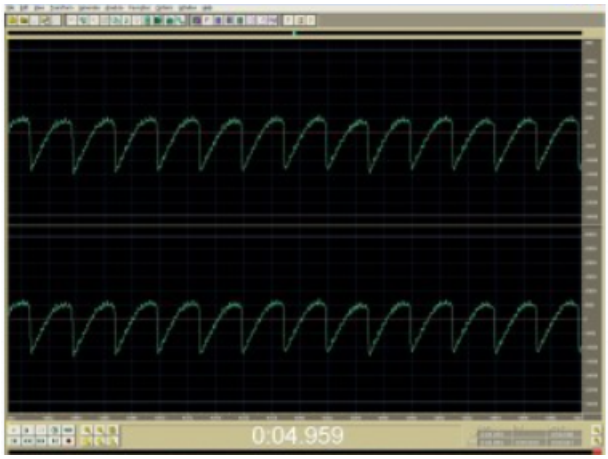
مبانی نظری موسیقی سنتی ایرانی

صوت نتیجه ارتعاش است و در محیط مادی مانند هوا یا آب به صورت موج انتشار می‌یابد و ما در دستگاه شنوایی‌مان آن را درک می‌کنیم. مفاهیمی چون شدت، بلندی، طول موج، بسامد یا فرکانس برای صدا قابل تعریف هستند و هنگامی که سیمی از ساز مرتعش و یک نت نواخته می‌شود، در واقع فرکانس ارتعاش سیم تغییر می‌کند و این تغییر در محیط هم منتقل می‌شود؛ فرکانس به ارتعاش درآوردن استخوان داخل گوش نیز، دقیقاً معادل همین فرکانس است (تعداد چرخه‌های موجی که در یک ثانیه اتفاق می‌افتد به عنوان فرکانس موج اندازه‌گیری می‌شود. صداهایی که توسط انسان شنیده می‌شود از ۲۰ هرتز تا ۲۰۰۰۰ هرتز است).

شکل امواج با وجود نواختن نت یکسان در یک گام، بسیار متفاوت است. به طور مثال، شکل‌های ۱.۲ ، ۲.۲ ۳.۲ و ۴.۲ همگی شکل امواج تولید شده توسط نتِ «می» در سازهای مختلف را نشان می‌دهند (برای رسم این اشکال از نرم‌افزار Cool Edit اَستفاده شده است). بنابراین آنچه موجب تشخیص اصوات از یکدیگر می گردد، شکل موج صداهای شنیده شده است. عوامل موثر در شکل موج بسیار پیچیده و زیاد هستند. مثلا آنچه در موسیقی به رنگ صدا معروف است، به عاملی به نام جمع هارمونی‌ها و شکل موج صوت مربوط می‌شود و همین تفاوت‌ها می‌تواند عامل تشخیص صدا باشد.

حال با توجه به تنوع بالای سازهای مورد استفاده در موسیقی سنتی ایرانی که بعضا در نواحی مختلف کشور با کوک و سبک‌ها و ساختار کمی متفاوت نواخته می‌شوند، گاهی صداهای تولید شده بسیار مشابه بوده و خصوصا در آثاری که به صورت گروهی نواخته می‌شوند، تشخیص تعداد و انواع سازها کاری بسیار دشوار است و برای افرادی که با موسیقی و تفاوت سازها آشنا نباشند، تشخیص آن‌ها تقریبا غیر ممکن بوده و با درصد خطای بسیار بالایی همراه است.

از این رو، در ادامه این بخش به بررسی مفاهیم و تعاریف موسیقی سنتی ایرانی پرداخته شده است.



شکل ۱.۲: شکل موج نت می در ویالون

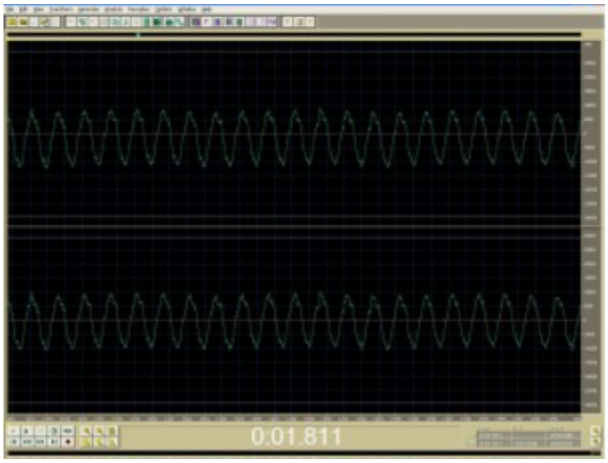
۱.۲ موسیقی ردیف دستگاهیِ ایران

ردیف

واژه ردیف در زبان فارسی به معنای راسته است. ردیف مجموعه‌ای از ملودی‌های سنتی ایرانی است که در قالب‌های مختلف دستگاهی و آوازی گردآوری شده و با نظم خاصی در قالب‌های دستگاه‌ها و آوازها تنظیم شده است. به عبارت دیگر ردیف روشی است برای طبقه‌بندی گوشه‌ها با هدف سهولت در آموزش با تکرار. نورعلی برومند، موسیقی‌دان ایرانی و نظریه‌پرداز ردیف، توضیح می‌دهد که ردیف، نماد اصلی و قلب موسیقی ایرانی است. او می‌افزاید که در مقایسه با سایر فرهنگ‌های موسیقی قومی، داشتن ردیف کاملاً منحصر به فرد است. بیشتر این ملودی‌های سنتی از منابع عامیانه و مردمی مشتق شده‌اند و هیچ سرنخ روشنی در مورد منشاء آنها وجود ندارد[۲۱].

دستگاه

دستگاه به چرخه‌ای چندوجهی گفته می‌شود که با یک مقام مادر شروع می‌شود و پس از به کارگیری تمام ظرفیت‌های این مقام اصلی، از مقام‌های دیگر برای ادامه روند موسیقی استفاده می‌کند. اما هر بار برای حفظ وحدت عمومی به مقام‌های اصلی باز می گردد. موسیقی ایرانی شامل هفت دستگاه اصلی و پنج دستگاه فرعی است که آواز نام دارند. واژه دستگاه به جای مقام در موسیقی ایرانی، به احتمال زیاد از دوره قاجاریه به بعد به کار رفته است. دستگاه ترکیبی از دو کلمه است؛ «دست» و «گاه». گاه در موسیقی کهن ایرانی به جای انگشتان روی ساز اطلاق می‌شود. بر این اساس، «دستگاه» به معنای نحوه قرار دادن دست‌ها (انگشت‌ها) روی ساز است و این حالت‌ها در هر



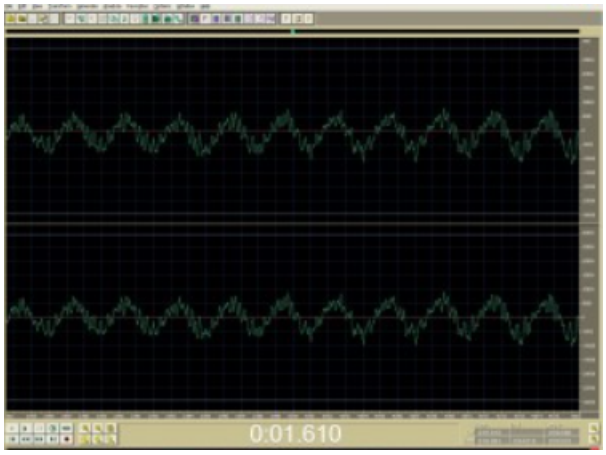
شکل ۲.۲: شکل موج نت می در پیانو

مقامی متفاوت است. آوازها نیز به عنوان دستگاه فرعی و مشتق، فقط با ساختارهای ساده‌تر، در نظر گرفته می‌شوند[۲۱].

همانطور که گفته شد، دستگاه یک نظام موسیقایی است که با یک مقام مادر شروع می‌شود و پس از به کارگیری تمام ظرفیت‌های این مقام پایه، از سایر مقام‌ها برای ادامه روند موسیقی استفاده می‌کند و هر بار به مقام اصلی باز می‌گردد. دستگاه از این نظر مجموعه‌ای از ملودی‌ها با کیفیت نیمه‌باز و نیمه‌بسته است. نیمه‌بسته به این معنا که رویدادهای موسیقایی خاص به طور قطع در دستگاه خاصی اتفاق می‌افتد و نیمه‌باز بدین معنا که اجراکننده موسیقی در اجرای عملی این رویدادها نوعی آزادی دارد.

در کل می‌توان گفت که موسیقی ایرانی فراز و نشیب‌های فراوانی داشته است؛ اما از دوره قاجار، موسیقی ایرانی وارد دسته‌بندی، تنظیم و تدوین تازه‌ای شده است که این تقسیم‌بندی امروزه نیز رواج دارد. این دسته‌بندی شامل هفت دستگاه، به نام های شور، سه گاه، چهارگاه، ماهور، همایون، نوا و راست پنج گاه، و پنج آواز می‌شود. از دستگاه شور، چهار شعبه یا آواز به نام‌های ابوعطا، افشاری، بیات ترك و دشتی منشعب می‌شود که به «متعلقات شور» موسوم اند. از دستگاه همایون هم آواز بیات اصفهان منشعب می‌شود. هر یک از این دوازده دستگاه و آواز موسیقی ایرانی، داراي گوشه‌های متعددی است. ازاین رو، هفت دستگاه موسیقی ایرانی، اساس و پایه موسیقی در ایران بوده و می‌توان آن را موسیقی رسمی کشور دانست.

از طرف دیگر، درست شبیه زبان فارسی که در کنار زبان‌ها و گویش‌های محلی، زبان رسمی مردم ایران است و تقریباً همه، در کنار گویش محلی خود، به این زبان رسمی نیز صحبت می‌کنند و آن را می‌فهمند، موسیقی دستگاهی و سازهایش نیز در تمام نقاط کشور به عنوان موسیقی رسمی و ملی ایران شناخته می‌شود. آنچه امروز به عنوان موسیقی کلاسیک یا رسمی ایران شناخته می‌شود،



شکل ۳.۲: شکل موج نت می در گیتار الکتریک

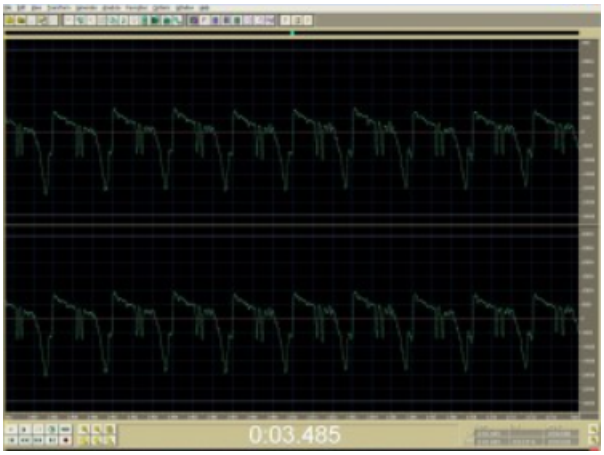
تاثیراتی از موسیقی نواحی و گویش‌های مختلف ایران گرفته است، اما این تاثیرات همه در قالب لهجه فارسی تغییر یافته و یا تعدیل شده است. امروزه، چکیده و برآیند فرهنگی این نوع موسیقی، در قالب ردیف موسیقی دستگاهی تجلی یافته است.

۲.۲ ارکان موسیقی

تفکیک مرز بین موسیقی سنتی ایرانی با دیگر انواع موسیقی (بخصوص موسیقی کلاسیک غربی که بیشتر پژوهش‌ها در این حوزه انجام شده)، ضروری به نظر می‌رسد. با این وجود، به طور کامل و قطعی نمی‌توان مبانی نظری موسیقی شرقی (و از جمله موسیقی سنتی ایرانی) را از مبانی نظری موسیقی غربی تفکیک نمود و برخی مباحث و ارکان، در هر دو نوع موسیقی، مشترک است. اصوات موسیقی، به واسطه چهار رکن اصلی و بنیادی از دیگر صداها متمایز می‌شوند: نواک، کشش (ارزش زمانی نغمه، دیرند، فاصله زمانی، مدت تداوم صوت، ریتم یا وزن)، دینامیک (شدت صوت یا دامنه) و طنین (رنگ صوتی یا شیوش)[۲۵].

نواک

نواک که در متون مختلف، با عناوین دیگری چون زیروبمی، پیچ فرکانسی، ارتفاع صوت یا نام نغمه نیز بکار رفته است، صفتی ادراکی از صداست که از طریق تنظیم فرکانس یک موج سینوسی، با دامنه دلخواه برای انطباق با صدای مورد نظر به دست می‌آید. ازاین رو، نواک هر نغمه را می‌توان در سه مفهومِ بسامد، نام نغمه و طولِ موج نیز توصیف کرد. هر چه طول موج کوتاه‌تر باشد، صدای حاصل زیرترِ خواهد بود و برعکس. الفبای موسیقی، نغمه نام دارد. برای همه صداهای موسیقایی، از



شکل ۴.۲: شکل موج نت می در آکاردئون

بم‌ترین تا زیرترین آن‌ها، فقط هفت نام وجود دارد که حداکثر نُه بار تکرار می‌شوند. نام‌گذاری این هفت نغمه به دو شکل هجایی یا الفبایی رایج است. نام گذاری نغمه‌ها در برخی کشورها (ازجمله فرانسه، ایتالیا و ایران) از نظام هجایی و در کشورهای انگلیسی زبان (انگلستان و امریکا) و آلمانی زبان، (اتریش و آلمان)، از نظام الفبایی تبعیت می کنند. از بین سه ویژگیِ نواک، ریتم و طنین، تنها ویژگی نواک در تعیین دستگاه‌های موسیقی ایرانی نقش دارد. ازاین رو، مهم‌ترین پردازشی که بر روی داده‌های خام باید انجام داد، به دست آوردن طیف فرکانسی و پیدا کردن فرکانس‌های غالب قطعه است. به نظر می رسد بیشتر پژوهش‌های انجام شده در این زمینه نیز به این مسئله واقف بوده و در اولین مرحله از شناسایی دستگاه، یعنی آشکارسازی و تشخیص نغمه‌ها، تنها از همین ویژگی استفاده کرده‌اند[۲۵].

۳.۲ کارکرد درجات در موسیقی غربی و ایرانی

گام دیاتونیک و همین طور گام‌های هفت دستگاه موسیقی سنتی ایرانی، متشکل از هشت نغمه پیاپی است که هر نغمه را درجه آن می‌نامند. درجه یکم را پایه یا تونیک، درجه دوم را روپایه، درجه سوم را میانی، درجه چهارم را زیرنمایان، درجه پنجم را نَمایان، درجه ششم را رونمایان، درجه هفتم را محسوس (حساس) و بالاخره درجه هشتم را هنگام (اُکتاو) نامند. هر درجه نام خاص، نقش و کارکرد ویژه‌ای در یک گام را بر عهده دارد؛ به عنوان مثال، درجه پنجم هر گام (نمایان) پس از پایه، مهم‌ترین نغمه گام است. همچنین، نغمه هفتم گام (محسوس)، واجد چنان حساسیتی است که موجب حرکت آن به طرف نغمه هنگام (یا پایه) می‌شود[۲۵].

۴.۲ انواع گام

گام بالقوه و بالفعل

یکی از ارکان تشخیص دستگاه‌ها و گوشه‌های موسیقی ایرانی، مبحث گام و تشخیصِ «فواصل بینِ نغمه‌ها» در این دستگاه‌ها است که در متون مختلف از آن با عنوانِ «شابلون»، «الگو» و «اِشِل صوتی» نیز یاد شده است. هرگاه یک الگوی موسیقایی در فاصله یک اُکتاو نوشته و استفاده شود، یک گام موسیقایی تشکیل می‌شود. به عبارت دیگر، گام از اصوات معین و پی درپی تشکیل شده که با فاصله‌های معین و حساب شده به دنبال هم قرار می‌گیرند و آخرین نغمه آن، اُکتاوِ بالایی است. آنچه در هر گام اهمیت دارد، تعداد اصوات پیاپی و نیز فاصله میان آن‌ها است که این عوامل، نوع گام را تعیین می‌کند. امروزه در موسیقی غربی، گام‌ها از هر نغمه که آغاز شوند، به دو گروه کلی تقسیم می شوند:

۱. گام‌های بالقوه

۲. گام‌های بالفعل

گام بالقوه یا کروماتیک، پایه و اساس تشکیل گام بالفعل یا دیاتونیک است. وقتی تمام نغمه‌های مورد استفاده در یک موسیقی، به طور مثال موسیقی ایرانی را به ترتیب از بم به زیر، در فاصله یک اُکتاو می‌نویسیم، به گام بالقوه آن موسیقی دست پیدا می‌کنیم. به عبارت دیگر، این گام با آغاز از هر نغمه‌ای، به فاصله‌های نیم‌پرده‌ای (در موسیقی غربی) یا ربع‌پرده‌ای (در موسیقی ایرانی) درجه‌ها را طی کرده، به تدریج بالا می‌رود تا در نغمه سیزدهم (در موسیقی کلاسیک غربی) و در نغمه هفدهم یا بیست و چهارم (در موسیقی ایرانی) به اُکتاو همان نغمه آغاز برسد.

فاصله بین درجه‌ها در گام بالفعل یا دیاتونیک، برخلاف گام کروماتیک بالقوه، یکدست و برابر نیست و الگوی قرار گرفتن این فاصله‌ها در هر گونه با گونه دیگر متفاوت است. امروزه در موسیقی کلاسیک غربی، رایج‌ترین انواع گام دیاتونیک (که به آن گام طبیعی نیز می‌گویند)، ماژور و مینور نام دارند که هر دو گام، شامل هشت نغمه بوده و مجموع فاصله‌های میان نغمه‌ها، متشکل از پنج پرده و دو نیم‌پرده است. گام بالفعل در موسیقی سنتی ایرانی به دستگاه تعبیر شده است. گام بالفعل، اساساً از ترکیب دو دانگ و یک فاصله طنینی تشکیل شده است. یکی از تفاوت‌های عمده گام بالفعل در موسیقی کلاسیک غربی و موسیقی ایرانی، نحوه قرار گرفتن دو دانگ است، به طوری که در موسیقی کلاسیک غربی، فاصله طنینی بینِ دو دانگ قرار می‌گیرد، درحالی که دستگاه‌ها و مُدهای اصلی در موسیقی ایرانی، بر اساس ترکیب دو دانگ پشت سرهم و پیوسته که در انتهای این دو دانگ، یک فاصله طنینی قرار می‌گیرد ساخته شده‌اند[۲۵].

۵.۲ فواصل موسیقایی

مفهوم فواصل در موسیقی ایرانی و اثرات آن بر ساختار و طبقه بندی ردیف موسیقی ایرانی، یکی از مهم‌ترین مسائلی است که مورد توجه موسیقی‌دان‌ها بوده است. اختلاف در نواک دو نغمه، فاصله موسیقایی را ایجاد می‌کند. انواع فواصل موسیقایی عبارت اند از: فواصل دوم (نغمات هم جوار)، فاصله سوم (فاصله شخصیت بخش)، فواصل درست (شامل فاصله هشتم یا اُکتاو، فاصله پنجم یا پنتاکورد و فاصله چهارم یا دانگ) و نیز فاصله ششم و هفتم. در ادامه، فاصله دوم و چهارم (دانگ) به دلیل اهمیت بیشتر توضیح داده شده اند[۲۴]:

نیم‌پرده در موسیقی غربی، کوچک‌ترین فاصله میان دو نغمه پیاپی است. این فاصله در موسیقی ایرانی ربع‌پرده است؛ با این تفاوت که این فاصله در موسیقی غربی تعدیل شده است ولی این امر در موسیقی ایرانی (و به طور گسترده تر در موسیقی مشرق زمین) به دلیل ماهیت این نوع موسیقی و ترجیح اساتید اهل فن انجام نگرفته است تا از این طریق، دست نوازنده را در اعمال سلایق خود در کوک فواصل ربع‌پرده و بداهه نوازی باز بگذارد.

در موسیقی ایرانی، به این دلیل که نغمات در اکثر موارد به صورت پیوسته و پلکانی حرکت می‌کنند، فاصله دوم رایج‌ترین فاصله‌ای است که در این موسیقی از آن استفاده می‌شود. با وجود دامنه وسیعی از انعطاف‌ها و کم وزیاد شدن‌های دو نغمه همسایه، از لحاظ نظری تمام گونه‌های فاصله دوم در موسیقی سنتی ایرانی، به یکی از انواع زیر تعلق دارند:

۱. دومِ کوچک (بقیه): ۱/۲ پرده

۲. دومِ خنثی (مجنب): ۳/۴ پرده

۳. دومِ بزرگ (طنینی): ۱ پرده

۴. دومِ بیش بزرگ (بیش طنینی): ۵/۴ پرده

در موسیقی ایرانی فاصله چهارم که به آن دانگ گفته می‌شود، از جایگاه ویژه‌ای برخوردار است. این فاصله از آن‌جا که فاصله‌ای «درست» است، به مثابه چارچوب عمل می‌کند و بین سه فاصله «درست» مهم (یعنی فواصل اُکتاو، پنجم و چهارم که هر سه فاصله از قدیم در ایران مرسوم بوده است) کوچک‌ترین فاصله است[۲۵]. گردش نغمات در موسیقی سنتی ایرانی عموماً در محدوده یک دانگ صورت می‌گیرند[۲۴]، بر اساس تجربه و تحلیل‌هایی که طی سالیان متمادی از رپرتوار موسیقی ایرانی انجام داده، به این نتیجه رسیده است که حتی پیچیده‌ترین ساختارهای مُدال در ردیف موسیقی ایرانی، بر اساس چهار دانگ بنیادین بنا شده‌اند(۱.۲).

جدول ۱.۲: چهار دانگ بنیادین و فواصل بین آنها از نغمهٔ سل تا نغمهٔ دو (از راست به چپ)[۲۵]

نام دانگ بنیادین	درجه اول	درجه دوم	درجه سوم	درجه چهارم
شور (S)	G	A	Bb	C
چهارگاه (C)	G	A	B	C
ماهور (M)	G	A	B	C
نوا (N)	G	A	Bb	C

در جدول ۲.۲، این دانگ‌ها، بر اساس فواصل بینِ درجه‌ها نشان داده شده‌اند. همان طور که گفته شد، گامِ بالفعل، اساساً از ترکیبِ دو دانگ و یک فاصله طنینی (دوم بزرگ) تشکیل شده است.

جدول ۲.۲: فواصل چهار دانگ بنیادین در موسیقی سنتی ایرانی (از چپ به راست)[۲۵]

نام دانگ بنیادین	فاصله اول	فاصله دوم	فاصله سوم
شور (S)	۳/۴	۳/۴	۱
چهارگاه (C)	۳/۴	۵/۴	۱/۲
ماهور (M)	۱	۱	۱/۲
نوا (N)	۱	۱/۲	۱

۶.۲ گوشه

هر گوشه شخصیت مستقل دارد. مدت زمان اجرای گوشه‌ها به ندرت از پنج دقیقه تجاوز می‌کند و بیشتر آن ها نسبتاً کوتاه هستند. مثلاً برخی گوشه‌های کوتاه نیز بین ۱۰ تا ۱۲ ثانیه اجرا می‌شوند. در تمام ردیف‌های موجود، برخی گوشه‌های دستگاه شور، مانند رضوی، شهناز و سلمک بلند و برخی دیگر، مانند زیرکش سلمک و گلریز کوتاه هستند. تبیین فرم واحد برای همه گوشه‌ها کار ساده‌ای نیست، ولی همان طور که در متن ادبی و یا درباره قالب شعری، مانند «نَحو» می‌توان راجع به غزل صحبت کرد، می‌توان راجع به آناتومی گوشه هم صحبت کرد و اجزاء و کارکرد قسمت هاي مختلف آن را مشخص کرد.

بر اساس تحلیل طلایی[۲۴]، از ردیف موسیقی سنتی ایرانی، یک گوشه کامل را می‌توان به ترتیب به جمله‌ها و بخش‌های: آغازین، معرّف، گسترشی، تکمیلی، پایانی و ختم تجزیه و تفکیک کرد . در این تبیین از آناتومی گوشه، جمله معرّف مهم‌ترین بخش گوشه بوده و شامل ملودی خاص و

یا دیگر ویژگی‌های منحصربه فردی است که مختص همان گوشه است و به همین جهت، گوشه با آن شناسایی می‌شود . این ویژگی انحصاری می‌تواند الگویی ریتمیک باشد، مثل کرشمه؛ یا ملودی‌های خاصی باشد، مثل لیلی و مجنون یا غم انگیز و یا می‌تواند تحریر خاصی باشد، مثل جوادخانی و یا پاساژ، مانند مُحیّر[۲۵].

۷.۲ کانال

کانال نمایش صدایی است که از یک نقطه می‌آید یا به آن می‌رود. یک میکروفون می‌تواند یک کانال صدا تولید کند و یک بلندگو می‌تواند برای مثال یک کانال صدا را بپذیرد. یک فایل صوتی دیجیتال می‌تواند حاوی چندین کانال داده باشد. موسیقی که برای گوش دادن به هدفون ترکیب می‌شود، به عنوان یک فایل با دو کانال ذخیره می‌شود؛ یکی به گوش چپ و دیگری به سمت راست ارسال می‌شود، در حالی که صدای فیلم با صدای فراگیر اغلب برای ۶ کانال ترکیب می‌شود.

فصل ۳

مرور کارهای دیگران

با توجه به پیشرفت در یادگیری با نظارت، اکنون می‌توان مدل‌هایی را آموزش داد که قادر به انجام موفقیت آمیز انواع وظایف صوتی باشند. با وجود موفقیت غیرقابل انکار، این رویکرد از نقوص عمده رنج می‌برد و یادگیری خودنظارتی تلاش می‌کند تا بر این محدودیت‌ها غلبه کند؛ با ایجاد امکان یادگیری از دادگان‌های بدون برچسبِ به طور گسترده در دسترس و با یادگیری نمایش‌هایی با هدف کلی که می‌توانند برای وظایف مختلف پایین‌دستی مورد استفاده مجدد قرار گیرند. به طور خلاصه، این رویکرد یک وظیفه کمکی را بر اساس داده‌های بدون برچسب موجود فرموله می‌کند و یک مدل کاملاً با نظارت برای حل این کار آموزش داده می‌شود. ایده اصلی این است که با حل وظیفه کمکی، مدل همچنین در حال یادگیری برخی نمایش‌های هدف کلی در فضا با ابعاد پایین‌تر است.

به عنوان مثال، رمزگذار، بخشی از معماری مدل که داده‌های ورودی را به فضای جاسازی نگاشت می‌کند، می‌تواند به عنوان استخراج‌کننده ویژگی برای وظایف مختلف پایین‌دستی مورد استفاده مجدد قرار گیرد. یکی از اولین موفقیت‌های یادگیری خودنظارتی، در زمینه مدل‌های زبانی به دست آمد و Word2Vec برای نگاشت کلمات با کدگذاری یک‌طرفه معرفی و استفاده شد.

۱.۳ اهمیت پژوهش در حوزه موسیقی سنتی ایرانی

حال با توجه به توضیحات و تعریفات مذکور، می‌توان به بررسی اهمیت و چگونگی بازیابی اطلاعات موسیقی سنتی ایرانی پرداخت. بیشتر موضوعات پژوهشی در زمینه بازیابی اطلاعات موسیقی و سایر حوزه‌های پژوهشی مرتبط با موسیقی، در مورد موسیقی کلاسیک و پاپ غربی است و پژوهش‌های منتشرشده در حوزه پردازش رایانه‌ای موسیقی سنتی ایرانی بسیار ناچیز است و به نظر می‌رسد دلیل این امر، گنگ و مبهم بودن مسئله و یا ناآشنایی با موسیقی سنتی ایرانی باشد. به عبارت دیگر، این امر را می‌توان ناشی از شناخته نشدن موسیقی ایرانی به صورت گسترده و خارج از ایران دانست؛ به

نحوی که نیاز به انجام چنین پژوهش‌هایی، براي پژوهشگران غیر ایرانی ایجاد نشده است. در ادامه این فصل به شرح معدود پژوهش‌های منتشرشده در این حوزه، پرداخته شده است.

در چند دهه اخیر، انواع دیگر موسیقی جهان مانند فرهنگ موسیقی هندی، ترکی و عربی نیز تا حدودی مورد مطالعه قرار گرفته است. در حالی که موسیقی ایرانی که یکی از مهم‌ترین سیستم‌های موسیقی خاورمیانه است و بسیاری از فرهنگ‌های موسیقی دیگر مانند ترکی، آذری و عربی ریشه در آن دارند، به شدت ناشناخته مانده است، زیرا تحقیقات بسیار محدودی در مورد آن انجام شده است. از طرفی پژوهش در این زمینه، با چالش‌های متفاوتی از موسیقی غرب روبرو است. به عنوان مثال، یکی از این چالش‌ها کار با موسیقی قومی است. اول اینکه هیچ نظریه استانداردی برای این فرهنگ‌های موسیقی وجود ندارد. به عنوان دلیل دوم نیز می‌توان به این نکته اشاره کرد که کار با موسیقی قومی در مقایسه با موسیقی غربی نیاز به رویکردهای جدیدی دارد، زیرا باید با گروه وسیعی از موسیقی، فرهنگ‌ها و نظریه‌ها کاملاً متفاوت با استانداردهای از پیش تعریف شده غربی کار کرد.

تحلیل نواک، یک موضوع مهم در زمینه بازیابی اطلاعات موسیقی است. تجزیه و تحلیل نواک در فرهنگ‌های موسیقی قومی مختلف نشان می‌دهد که مفهوم غربی نواک (دوازده کلاس استاندارد صدای غربی) در بسیاری از فرهنگ‌های موسیقی غیرغربی دیگر بسیار محدود و یا حتی بی‌معنی است، زیرا موسیقی همیشه از دسته‌های نواک مجزا تشکیل نمی‌شود و مقام و دستگاه در فرهنگ‌های موسیقی خاورمیانه نمونه‌های نقض خوبی هستند. در فرهنگ‌های موسیقی مختلف ساختارهای فاصله‌ای متفاوتی وجود داشته و دارد و این ساختارها در طول تاریخ بارها تغییر کرده‌اند. تفاوت در مفهوم نواک در فرهنگ‌های مختلف، تنها گوشه‌ای از تفاوت‌های موسیقی غرب و شرق است. از این رو، اهمیت پژوهش در حوزه موسیقی خاورمیانه، به خصوص موسیقی سنتی ایرانی بسیار زیاد است[۲۱].

با توجه به مطالبی که در بالا ذکر شد، تحقیق در مورد فرکانس‌ها و نواک‌ها در موسیقی فرهنگ‌های اقوام مختلف کار بسیار دشواری است و نرم‌افزارها، روش‌ها و الگوریتم‌هایی که بر اساس موسیقی کلاسیک غربی تعریف شده‌اند، برای این منظور مفید نیستند. در استفاده از برنامه‌ها و نرم‌افزارهای موسیقی قومی و جهانی که در اصل برای موسیقی غربی ساخته شده‌اند، مشکلات متعددی وجود دارد. مشکل اول به دلیل فضاهای نواک متفاوت در موسیقی غربی و سایر فرهنگ‌های موسیقی مانند آسیایی، آفریقایی، هندی و خاورمیانه است. مورد دوم فقدان تئوری موسیقی برای موسیقی قومی و غیرغربی است.

۲.۳ پژوهش‌های مرتبط با موسیقی غیر ایرانی

با هدف توسعه نرم‌افزار، برنامه‌ها و فناوری‌های موسیقی غیرغربی، گروه فناوری موسیقی دانشگاه پومپئو فابرا در بارسلونا با همکاری شورای تحقیقات اروپا، یک پروژه تحقیقاتی به نام Music Comp را انجام داد که از سال ۲۰۱۱ آغاز و تا ۲۰۱۷ ادامه داشت. این پروژه در رابطه با

پردازش اطلاعات موسیقی انجام شد که در آن پنج فرهنگ موسیقی سنتی نقاط مختلف جهان از جمله هندوستانی، موسیقی شمال هند، کارناتیک، موسیقی جنوب هند، مقام، سیستم موسیقی ترکیه، عربی_اندلس، موسیقی شمال غرب آفریقا و اپرای پکن، موسیقی سنتی چین، حوزه تمرکز بود. نرم افزاری مانند Dunya یک برنامه برای دسترسی و تجسم داده‌های موسیقی مانند ضبط‌های صوتی، ویژگی‌های استخراج شده و نتایج تجزیه و تحلیل است. برنامه Saraga یک برنامه اندرویدی برای موسیقی کارناتیک و هندوستانی، PycompMusic ، مرورگر Dynya ، جعبه ابزار تحلیل موسیقی مقام ترکی_عثمانی و برخی برنامه ها و نرم افزارهای دیگر، همگی دستاوردهای پژوهش در این زمینه هستند[۲۱].

روش های یادگیری با نظارت به طور گسترده در کارهای موسیقی غیر ایرانی مانند تشخیص آکورد، تشخیص کلید، ردیابی ضربات، برچسب گذاری صوتی موسیقی و توصیه‌گرهای موسیقی استفاده شده است.

کدگذاری پیش‌بینی متضاد، یک رویکرد جهانی برای یادگیری متضاد است، و برای دسته‌بندی گوینده و واج با استفاده از اصوات خام، در میان کارهای دیگر، موفق بوده است. در [۱۷] چندین عملگر خودنظارتی برای حل وظایف رگرسیون یا تبعیض باینری، که به طور مشترک یک رمزگذار را برای تشخیص گفتار بهینه می کنند، معرفی شده است. برای بهبود نمایش‌های شرایط آکوستیک ناسازگار و قابلیت انتقال آن‌ها، تقویت‌کننده‌ها را به سیگنال گفتار ورودی اعمال می‌کنند [۱۸]. در بازیابی اطلاعات موسیقی، پیشرفت‌های اخیر در تخمین نواک خودنظارت شده [۱۱] صورت گرفته است.

همچنین Audio2Vec در حوزه فرکانس زمانی عمل می‌کند و با بازسازی برش‌های طیف‌گرام از برش‌های گذشته و آینده آموزش می‌بیند[۲۳]. Audio2Vec با داده‌های محدود، از مدل‌های با نظارت در دسته‌بندی نواک و ساز بهتر عمل می‌کند. CLAR همچنین با بهره‌گیری از یادگیری متضاد، زیان را در ترکیبی از نمایش‌های آموخته‌شده از صداهای خام و طیف‌نگارهای مل محاسبه می‌کند [؟]. COLA از روش مشابهی اما فقط با طیف‌نگارهای مل، استفاده می‌کند و از مقایسه‌های دوخطی به جای شباهت کسینوس استفاده می‌کند [۲۰]. هر دو کار، بر اساس دستور گفتار، دسته‌بندی صدای محیطی، و دسته‌بندی نواک و ساز با استفاده از دادگان NSynth [۹] ارزیابی می‌شوند.

۳.۳ پژوهش‌های مرتبط با موسیقی سنتی ایرانی

بطور کلی، سیگنال موسیقی به عنوان یک پدیده پیچیده حاوی حجم زیاد و متنوعی از اطلاعات در خصوص ژانر، احساس، هنرمند، ساز و غیره است. تنوع بالای اطلاعات موجود در سیگنال موسیقی، باعث مطرح شدن حیطه گسترده‌ای از مسائل در بازیابی اطلاعات موسیقی مبتنی بر محتوا جهت مطالعه و پژوهش می‌شود که برخی از این مسائل عبارتند از: قطعه‌بندی یک قطعه موسیقی به بخش‌های آواز و غیرآواز، شناسایی خواننده، دسته‌بندی ژانر، جستجو با زمزمه، تشخیص بار احساسی موسیقی، تشخیص ساز موسیقی، حاشیه‌نویسی خودکار موسیقی و غیره[۱].

۱.۳.۳ تشخیص دستگاه

همان طور که گفته شد، تعداد کل پژوهش‌های انجام شده که به طور خاص به دسته‌بندی دستگاه‌ها و گوشه‌های موسیقی سنتی ایرانی پرداخته‌اند، معدود است که یکی از آن‌ها[۱۳] تنها به موضوع تعیین فرکانسِ پایه نغمه در یک ساز ایرانی پرداخته و وارد مقوله تشخیص دستگاه و گوشه نشده است.

در دو دهه اخیر تحقیقاتی در زمینه اتنوموزیکولوژی محاسباتی با تمرکز بر موسیقی ایرانی انجام شده است. دارابی و همکاران [۸] تحقیقی به منظور شناخت دستگاه و مقام برای موسیقی ایرانی انجام داده‌اند که در آن به کارگیری روش‌های تشخیص الگو در شناسایی دستگاه‌ها و مقام‌ها در موسیقی ایرانی با توجه به حالت‌های موسیقی مورد بررسی قرار می‌گیرد و بر دستگاه همایون و تئوری تقسیم هر اکتاو به ۶۰ بازه تمرکز می‌کنند.

در [۲] از یک شبکه عصبی با توابع شعاعی پایه برای تشخیص دستگاه ماهور از سایر دستگاه‌ها برای ساز سه‌تار بهره برده است. دادگان استفاده شده شامل ۱۳۵ قطعه موسیقی است که ۶۰ تای آن در دستگاه ماهور و بقیه در پنج دستگاه دیگر بوده است. بعد از آموزش شبکه RBF برای ۷۰ درصد داده‌های موجود در دادگان، نهایتاً به دقت حدود ۷۳ درصد در تشخیص دستگاه ماهور رسیده است.

در پژوهشی دیگر [۳]، بر این اساس که نت‌های نواخته شده توسط ساز نقش کلیدی در تشخیص دستگاه‌های موسیقی ایفا میکند، سعی بر آن داشته که نت‌های قطعه موسیقی را با دقت بالایی استخراج نماید و در ادامه با مشخص کردن فواصل بین این نت‌ها و با توجه به منحصر به فرد بودن الگوهای این فواصل برای دستگاه‌های مختلف، دستگاه قطعه موسیقی مورد نظر را تشخیص دهد. در این پژوهش برای ارزیابی روش پیشنهادی، از ۴۶ قطعه موسیقی (۴۲ قطعه توسط ساز تار و ۴ قطعه توسط سنتور نواخته شده است) در پنج دستگاه مختلف استفاده شده و به دقت ۹۳ درصد رسیده است.

در پژوهش[۱۵]، براي دسته‌بندی ردیف میرزا عبدالله از ویژگی‌های مختلفی نظیر ناهمگونی، ضرایب کپسترال بر مبنای مقیاس مل، فرکانس گام، میانگین و انحراف معیار سنتروید طیفی بهره برده شده است. برای دسته‌بندی روش‌های مختلفی نظیر ماشین بردار پشتیبان شبکه عصبی پرسپترون و K تا نزدیکترین همسایه، آزموده شده است که ماشین بردار پشتینان به دقت بالاتری دست یافته است. دادگان استفاده شده شامل ۱۲۵۰ قطعه موسیقی از سازهای زهی زخمه‌ای تار و سه‌تار توسط چهار استاد معروف ایرانی است و دربرگیرنده هفت دستگاه و شش آواز است.

اخیراً در پژوهشی [۱۹] دیگر، دادگانی با عنوان دادگان موسیقی سنتی ایرانی مریم به منظور تشخیص دستگاه در موسیقی سنتی ایرانی به صورت مستقل از نوع ساز معرفی شده است. این دادگان شامل ۱۱۳۷ قطعه موسیقی است که ۶۳۱ تای آن صدای نی با صدای برخی سازهای دیگر در پس‌زمینه است و در بقیه قطعات، صدای ویولن به عنوان صدای پیش‌زمینه است. قطعات این دادگان در هفت دستگاه است و تعداد قطعات انتخاب شده در هر دستگاه متفاوت است. دستگاه شور با ۴۴۵ قطعه بیشترین و دستگاه سه‌گاه با ۷۴ قطعه کمترین تعداد قطعه در این دادگان را دارد.

محققین این پژوهش با انتخاب دو ساز نی و ویولن، سعی بر ارائه دادگانی براي مسئله تشخیص دستگاه به صورت مستقل از ساز داشته‌اند که انتخاب تنها دو ساز برای این ادعا مقعول به نظر نمیرسد. در ادامه، این محققین از ۸۰ درصد داده‌های هر دستگاه در دادگان مریم برای آموزش و از ۲۰ درصد باقیمانده برای آزمون آن استفاده کرده‌اند. ایشان با این ادعا که دستگاه بر مبنای ۱۶ ثانیه موسیقی قابل تشخیص است، هر قطعه موسیقی به قطعات ۱۶ ثانیه‌ای قطعه‌بندی شده است. بر روی هر قطعه ۱۶ ثانیه‌ای تبدیل فوریه زمان کوتاه ۱۳ اعمال شده و ویژگی‌های حاصل به یک شبکه عصبی ژرف)با معماري: پنج لایه پیچشی + دو لایه GRU + دو لایه اتصال کامل) خورانیده شده است و متوسط امتیاز F۱ حدود ۸۶ درصد بر روی هفت دستگاه گزارش شده است.

در [۴] برای تشخیص دستگاه، با بهره‌گیری از منطق و نظریه مجموعه‌های فازی و با این فرض که هر نت نواخته شده، یک مجموعه فازی است؛ هر قطعه موسیقی را مجموعه‌ای از مجموعه‌های فازی در نظر می‌گیرد و بر این اساس، به محاسبه شباهت بین دستگاه قطعه موسیقی ورودی و دستگاه‌های مرجع می‌پردازد دادگان استفاده شده شامل جمعاً ۲۱۰ قطعه موسیقی سنتی ایرانی با ۸۹ قطعه در دستگاه شور و نوا، ۳۰ قطعه در دستگاه سه‌گاه، ۴۱ قطعه در دستگاه ماهور و راست پنج‌گاه، ۲۶ قطعه در دستگاه همایون، و ۲۴ قطعه در دستگاه چهارگاه است. قطعات عمدتاً شامل آواز سه استاد آواز و تکنوازی چهار ساز تار، سه تار، سنتور و کمانچه است در این پژوهش نشان داده شده است که در روش پیشنهادی، یک دقیقه موسیقی از هر قطعه برای تشخیص دستگاه آن قطعه لازم و کافی است.

در [۵] براي دسته‌بندی هفت دستگاه موسیقی ایرانی، از شبکه عصبی پرسپترون چند لایه استفاده می‌شود. ورودی‌های شبکه عصبی، بیست قله بلندتر از طیف فرکانس هر قطعه موسیقی است. نتایج نشان می‌دهد شبکه می‌تواند دستگاه قطعات آزمون را با دقت حدود ۶۵ درصد برای نی، ۷۲ درصد برای ویولن و ۵۶ درصد برای آواز تشخیص دهد.

در [۱۲] نیز از یک شبکه عصبی پرسپترون با یک لایه مخفی برای دسته‌بندی پنج دستگاه شور، ماهور، همایون، سه‌گاه و چهارگاه استفاده شده است. ورودی شبکه، بردارهای باینری ۲۴ مولفه‌ای است که هر مولفه آن گویای یک نت در یک اکتاو است. شبکه به کمک ۱۲۰ الگوی آموزشی تولید شده، آموزش داده شده است و دقت ۱۰۰ درصد بر روی این ۱۲۰ الگوی ممکن گزارش شده است[۱].

عمده‌ترین محدودیت‌هایی که بر سر راه پژوهش‌های مرتبط با دسته‌بندی و تشخیص خودکار دستگاه‌های موسیقی ایرانی وجود داشته را می‌توان در سه مقوله کلی جای داد:

- انجام پژوهش‌ها به صورت موازی و مجزا
- نبود پایگاه داده منسجم
- نداشتنِ دانش کافی پژوهشگران از مبانی نظری موسیقی سنتی ایرانی

این عوامل در کنار یکدیگر، منجر شده تا نتایج پژوهش‌های انجام شده در این حوزه، عمدتاً از کارآمدی و مطلوبیت لازم برخوردار نبوده و به مرحله پیاده‌سازی و تولید نرم افزار کاربردي نرسد.

نبود یک پایگاه داده کامل و جامع، سبب شده تا مقایسه کارایی و علمکرد الگوریتم‌های پیشنهادی و تعمیم نتایج این نوع پژوهش‌ها امکان‌پذیر نباشد. در چنین شرایطی، گاه پژوهشگر در تجزیه و تحلیل داده‌ها با دشواری مواجه شده و از قابلیت اعتماد به نتایج پژوهش‌ها نیز کاسته خواهد شد. ازاین رو، ضروری است پایگاه داده‌ای ایجاد و ارائه شود که همه گوشه‌ها، دستگاه‌ها، ردیف‌ها و نیز قطعات مهم ساخته شده در موسیقی سنتی ایرانی را دربر گرفته و شامل نمونه‌ها و اجراهای مختلفی از کلیه سازهای ملی باشد[۲۵].

همانطوری که از مرور انجام شده در بالا قابل برداشت است، پژوهش‌های صورت گرفته در حوزه پردازش موسیقی سنتی بسیار ناچیز است. بخش عمده‌ای از این پژوهش‌ها بر روی مسئله تشخیص دستگاه متمرکز بوده است که دادگان‌های استفاده شده آن‌ها به لحاظ حجم، تنوع ساز و دستگاه از جامعیت کافی برخوردار نیستند.

فصل ۴

یادگیری خودنظارتی

روش اصلی و مورد نظر در این پروژه، روش یادگیری خودنظارتی است و با توجه به اهمیت این روش، در این فصل به بررسی آن پرداخته شده است.

۱.۴ مشکلات و محدودیت‌های یادگیری با نظارت

به لطف پیشرفت در یادگیری با نظارت، اکنون می‌توان مدل‌هایی را آموزش داد که قادر به انجام موفقیت‌آمیزِ انواع وظایف صوتی باشند. این روش‌ها به پیکره‌های برچسب‌دار نیاز دارند که همانطور که گفته شد، ایجاد آن‌ها در حوزه موسیقی دشوار، پرهزینه و زمان‌بر است، در حالی که داده‌های موسیقی بدون برچسب خام، در مقادیر زیادی موجود است. جایگزین‌های بی نظارت برای یادگیری عمیق، اگر بتوانند به دادگان‌های کوچکتر تعمیم داده شوند، جایگزین مناسبی برای این روش خواهند بود. علی‌رغم اهمیت یادگیری بی نظارت برای سیگنال‌های صوتی خام، یادگیری بی نظارت برای کارهای موسیقی هنوز پیشرفت‌هایی قابل مقایسه با یادگیری با نظارت نداشته است. موفقیت‌هایی با روش‌هایی مانند PCA و PMSC و K میانگین و اخیراً با روش‌های خودنظارتی در حوزه زمان- فرکانس برای وظایف دسته‌بندی کلی صدا، به دست آمده است. اما یادگیری بازنمایی‌های موثر از صدای خام به شیوه‌ای بی نظارت هنوز بسیار نوپا و جدید است. موفقیت مداوم روش‌های یادگیری عمیق، به کیفیت بازنمایی‌هایی بستگی دارد که به طور خودکار از داده‌ها کشف می‌شوند. این نمایش‌ها باید ساختارهای اساسی مهم را از ورودی خام، به عنوان مثال، مفاهیم میانی، ویژگی‌ها، یا متغیرهای پنهانی که برای وظایف پایین‌دستی مفید هستند، دریافت کنند. در حالی که یادگیری با نظارت با استفاده از مجموعه‌های بزرگ برچسب‌گذاری شده می‌تواند از نمایش‌های مفید استفاده کند.

همانطور که گفته شد، جمع آوری مقادیر زیادی از نمونه‌های برچسب‌گذاری شده پرهزینه و زمان‌بر است و همیشه امکان پذیر نیست. به طور خاص، در حوزه موسیقی، بسیاری از سازهای

کم منبع وجود دارند، که در آن‌ها پیشرفت به‌طور چشمگیری کندتر از سازهای با منابع بالا و رایج است. علاوه بر این، برچسب‌گذاری دادگان موسیقی، بسیار دشوار و پیچیده و نیازمند نیروی انسانی زیادی است (جمع متنوعی از موسیقیدان‌ها و نوازندگان) و به همین دلیل، امکان رخداد خطا هنگام برچسب‌گذاری بسیار بالاست. به طور مثال، در شبکه‌های مبتنی بر گرادیان، خروجی از طریق یک رویکرد پیش‌خور محاسبه می‌شود و برای محاسبه خطاها، به روش پس‌انتشار عمل می‌شود. این روش به داده‌های زیادی برای یادگیری نیاز دارد و همچنین زمان بیشتری برای یادگیری نیاز دارد (زیرا به صورت تکراری یاد می‌گیرد). علاوه بر این، زمانی که داده‌های جدید وارد می‌شوند، باید تمام پارامترها را مجدداً یاد بگیرد تا اطلاعات جدید را در خود جای دهد.

به طور کلی، یادگیری با نظارت، با مشکلات ذیل روبه‌رو است:

- برای انجام یادگیری با نظارت، به مقدار زیادی داده برچسب دار نیاز است.
- همچنین به مقدار زیادی داده نیاز است.
- جمع آوری داده‌ها پرهزینه است.
- برچسب زدن داده‌ها خسته کننده و گاهاً با خطای انسانی همراه است.
- برچسب گذاری داده‌ها اغلب به کمک متخصصان نیاز دارد.
- برچسب‌گذاری برای هر کار، به صورت جدا و اختصاصی انجام می‌شود. بنابراین برای تولید یک دادگان جدید برای هر کار جدید، نیاز به صرف وقت و هزینه است.
- روزانه حجم زیادی داده جدید تولید می‌شود که با این روش، همگی نیازمند به برچسب‌گذاری است.

راه‌هایی برای کاهش این محدودیت‌ها، یادگیری بی نظارت و خودنظارت است. به دنبال محبوبیت روزافزون، تلاش هایی برای گسترش یادگیری خودنظارتی برای کشف بازنمایی‌های صوتی و گفتاری انجام شده است و روش اصلی و مورد نظر در این پروژه نیز هست.

اخیراً، پیش‌آموزش شبکه‌های عصبی به‌عنوان یک روش مؤثر برای زمانی که داده‌های برچسب‌گذاری شده کمیاب هستند، پدیدار شده است. ایده کلیدی، یادگیری نمایش‌های کلی در حالاتی است که در آن مقادیر قابل‌توجهی از داده‌های برچسب‌دار یا بدون برچسب در دسترس است و در ادامه استفاده از نمایش‌های آموخته شده برای بهبود عملکرد در یک وظیفه پایین‌دستی با مقدار داده محدود است. این روش به ویژه برای کارهایی که در آن برای به دست آوردن داده‌های برچسب گذاری شده به تلاش قابل توجهی نیاز است، جالب توجه است.

۲.۴ مزایای یادگیری خودنظارتی

اگرچه انسان‌ها در درک و فهمیدن صداها مهارت دارند، ساخت الگوریتم‌ها برای انجام همین کارهای به ظاهر ساده انسانی برای ماشین، یک چالش بزرگ است؛ که این به دلیل وجود طیف وسیعی از تغییرات و ناپایداری‌ها در ویژگی‌های شنیداری است. دستیابی به ادراک شنیداری خودکار، مستلزم یادگیری بازنمایی‌های مؤثر است. اغلب پژوهش‌های انجام شده تا کنون، از طریق رویکردهای تمایزی، بازنمایی‌های مؤثری را به دست آورده‌اند؛ یعنی شبیه به رویکرد یادگیری با نظارت، مدل، نگاشت بین سیگنال ورودی به برچسب کلاس را می‌آموزد. فرض اساسی هنگام استفاده از چنین رویکردی این است که بازنمایی‌های نهفته، بازنمایی‌های مؤثری برای وظایف طراحی شده دارند. اما یکی از مشکلات اساسی این بازنمایی‌های آموخته شده، محدودیت بالقوه برای تعمیم‌پذیری است. به دلیل این که اولاً، این بازنمایش‌ها نیازمند و محدود به در دسترس بودن داده‌های برچسب‌گذاری‌شده و پرهزینه هستند. ثانیاً، بازنمایی‌ها به سمت یک حوزه خاص، مانند گفتار، موسیقی و غیره، منحرف می‌شوند.

بنابراین، در هر دو مورد، تنظیم دقیق و تعدیل داده‌های آموزشی هدفمند مورد نیاز است. روش جایگزین، رویکردهای اخیر یادگیری خودنظارتی با استفاده از یادگیری متضاد در فضای پنهان است. نشان داده شده که این رویکرد، بازنمایی‌های کارآمدی را یاد می‌گیرد که به عملکرد پیشرفته در تصاویر و فیلم‌ها به دست می‌آورد. یادگیری خودنظارتی به عبارتی، نوعی یادگیری بی نظارت است؛ زیرا از معیارهای مشابهی پیروی می‌کند و با داده‌هایی که هیچ برچسبی ندارند، سر و کار دارد. با این حال، به جای یافتن الگوهای سطح بالا و کلی برای خوشه‌بندی، یادگیری خودنظارتی همچنان تلاش می‌کند تا کارهایی را که به طور معمول توسط یادگیری با نظارت هدف قرار می‌گیرند، بدون هیچ گونه برچسب‌گذاری حل کند. به عبارت ساده‌تر، شکلی از یادگیری بی نظارت است که در آن داده‌ها، نظارت را فراهم می‌کنند. انگیزه پشت یادگیری خودنظارتی این است که ابتدا بازنمایی‌های مفید داده‌ها از مجموعه داده‌های بدون برچسب با استفاده از روش خودنظارتی آموخته شود و سپس نمایش‌ها با چند برچسب برای کاری با نظارت تنظیم شود. یادگیری با نظارت مورد نظر می‌تواند به سادگی دسته‌بندی تصویر یا تشخیص اشیا و غیره باشد. در یادگیری خودنظارتی، به طور کلی، بخشی از داده‌ها مخفی شده و شبکه موظف به پیش‌بینی آن بخش است؛ یعنی شبکه مجبور می‌شود آنچه را که واقعاً به آن اهمیت می‌دهیم (مثلاً یک نمایش معنایی)، بیاموزد.

یادگیری خودنظارتی سیستم‌های هوش مصنوعی را قادر می‌سازد تا داده‌های بیشتری را از مرتبه‌های بزرگی بیاموزند. یادگیری خودنظارتی مدت‌هاست که موفقیت زیادی در پیشرفت حوزه پردازش زبان طبیعی داشته است. سیستم‌هایی که به این روش از قبل آموزش داده شده‌اند، عملکرد بسیار بالاتری نسبت به زمانی دارند که صرفاً به روش با نظارت آموزش داده می‌شوند.

یادگیری خودنظارتی بر اساس یک شبکه عصبی مصنوعی است. شبکه عصبی در دو مرحله یاد می‌گیرد. اول، مسئله‌ای بر اساس شبه برچسب‌ها حل می‌شود که به مقداردهی اولیه وزن شبکه کمک می‌کند. سپس، مسئله واقعی با یادگیری با نظارت یا بی نظارت انجام می‌شود؛ به همین دلیل است

که می‌توان آن را به عنوان یک شکل میانی بین یادگیری با نظارت و بی نظارت در نظر گرفت.

۱.۲.۴ یادگیری بازنمایی خودنظارتی

یادگیری بازنمایی خودنظارتی، SSRL ، عمدتاً بر روی دقت در داده کم تمرکز دارد؛ با این حال از چندین مزیت بالقوه دیگر نیز برخوردار است[۱۰]:

۱. هزینه محاسباتی تنظیم یک مدل با نظارت از ابتدا، کمتر از آموزش از ابتدا است (اگرچه با تنظیم یک ویژگی از پیش‌آموزش دیده قابل مقایسه است).

۲. اگر وظیفه هدفِ با نظارت، با برچسب‌های پرت روبرو شود، آموزش در مقایسه با استفاده از برچسب‌های دقیق منجر به عملکرد بسیار بدتری می‌شود. با این حال SSRL انعطاف‌پذیری را در برابر چنین نویز برچسبی افزایش می‌دهد، که اغلب در عمل رخ می‌دهد.

۳. با توجه به یک سیستم آموزش‌دیده، SSRL همچنین می‌تواند استحکام تشخیص را در برابر حملات خصمانه، و همچنین خرابی‌های رایج مانند تاری (در بینایی) و نویز را بهبود بخشد.

۴. علاوه بر این، SSRL می‌تواند برای جلوگیری از پیش‌بینی‌های خودکار، یا تشخیص خارج از توزیع استفاده شود.

۳.۴ یادگیری خودنظارتی، یادگیری بازنمایی خودنظارتی و یادگیری نیمه‌نظارتی

در مواردی که مجموعه داده‌های منبع و هدف از نظر محتوا و فضای برچسب یکسان یا مشابه هستند، هر دو رویکرد نیمه‌نظارتی، SSRL ، و خودنظارتی، SSL ، به طور بالقوه می‌توانند اعمال شوند. از آنجایی که هر دو خانواده روش‌ها به سرعت در حال پیشرفت هستند و مقایسه مستقیم کمی وجود دارد، هنوز مشخص نیست که آیا یک خانواده باید ترجیح داده شود یا نه. با این حال، این دو استراتژی در اصل هر دو می‌توانند برای یک مشکل یادگیری اعمال شوند.

هنوز تحقیقات بسیار کمی در مورد میزانی که این استراتژی‌ها می‌توانند مکمل باشند و عملکرد را در صورت استفاده با هم افزایش دهند، انجام شده است. با این حال، نتایج اولیه در حوزه متن نشان می‌دهد که SSL و SSRL زمانی که با هم استفاده می‌شوند می توانند هم‌افزا باشند[۱۰]

همانطور که بیان شد، معمولاً این نوع یادگیری بر اساس شبکه‌های عصبی است. در یادگیری خودنظارتی، شبکه‌های عصبی می‌توانند در دو مرحله یاد بگیرند:

– برای مقداردهی اولیه وزن شبکه‌ها، مشکلات مربوط به برچسب‌های غلط را می‌توان حل کرد.

– وظیفه واقعی فرآیند را می‌توان با یادگیری با نظارت یا بی نظارت انجام داد.

یادگیری خودنظارتی، در سال های اخیر محبوبیت بیشتری پیدا کرده است و پیشرفت‌ها و نتایج امیدوارکننده‌ای را در سال‌های اخیر و در زمینه‌های مختلف ایجاد کرده است. یادگیری خودنظارتی، روشی را که انسان‌ها برای دسته‌بندی اشیاء یاد می‌گیرند، تقلید می‌کند.

وقتی از نتایج صحبت می‌شود، در سال‌های اخیر نتایج امیدوارکننده و دقیق مختلفی مشاهده شده است و شرکت‌های بزرگ مختلفی مانند «متا» و «گوگل» از این نوع فرآیند یادگیری برای پردازش تصویر، فیلم و صدا استفاده می‌کنند. ایده اصلی در پشت یادگیری خودنظارتی، آموزش الگوریتم‌ها با داده‌های با کیفیت پایین‌تر است؛ در حالی که سایر فرآیندهای یادگیری بر بهبود نتیجه نهایی الگوریتم‌ها متمرکز هستند.

همچنین این نکته شایان ذکر است که در عصر حاضر، میزان داده‌های تولید شده در حال افزایش است و پیچیدگی حاشیه‌نویسی داده‌ها نیز روز به روز افزایش می‌یابد. برای حل مسئله حاشیه‌نویسی، اینجا نیز روش های یادگیری خودنظارتی وارد تصویر می‌شوند. مدل‌های خودنظارت می‌توانند از داده‌های خام بهتر یاد بگیرند. در این پروژه به به نوعی از یادگیری خودنظارتی پرداخته شده که به یادگیری خودنظارتی متضاد معروف است. روش‌های متضاد خودنظارت، با یادگیری تفاوت‌ها یا شباهت‌های بین اشیا، بازنمایی‌هایی را ایجاد می‌کنند.

روش‌های یادگیری خودنظارت را می‌توان به دو دسته تقسیم کرد:

- یادگیری خودنظارتی متضاد
- یادگیری خودنظارتی غیر متضاد

در این پروژه تمرکز بر روی یادگیری خودنظارتی متضاد است.

۱.۳.۴ یادگیری خودنظارتی متضاد

موفقیت اخیر در مدل‌های خودنظارتی را می‌توان به علاقه مجدد محققان به کاوش در یادگیری متضاد نسبت داد. به عنوان مثال، انسان‌ها می‌توانند اشیاء را در طبیعت شناسایی کنند، حتی اگر به یاد نیاورند که جسم دقیقاً چه شکلی است و این کار با به خاطر سپردن ویژگی‌های سطح بالا و نادیده گرفتن جزئیات در سطح میکروسکوپی انجام می‌شود. بنابراین، اکنون سؤال این است که آیا می‌توان الگوریتم‌های یادگیری بازنمایی را طوری ساخت که بر جزئیات سطح پیکسل تمرکز نکنند و فقط ویژگی‌های سطح بالا را به اندازه کافی برای تشخیص اشیاء مختلف رمزگذاری کنند؟ با یادگیری متضاد، محققان در تلاش هستند تا به این موضوع بپردازند.

یادگیری متضاد برای یافتن داده‌های با کیفیت خوب استفاده می‌شود. می‌توان گفت یادگیری متضاد رویکردی برای یافتن اطلاعات مشابه و غیر مشابه از یک دادگان برای الگوریتم یادگیری

جدول ۱.۴: تفاوت یادگیری خودنظارتی متضاد و غیر متضاد

	یادگیری خودنظارتی متضاد	یادگیری خودنظارتی غیر متضاد
۱	از نمونه‌های مثبت و منفی از داده‌ها استفاده می‌کند.	فقط از نمونه‌های مثبت از داده‌ها استفاده می‌کند.
۲	فاصله بین نمونه‌های مثبت به حداقل می‌رسد.	روی کمینه مفید محلی از داده‌ها کار می‌کند.
۳	در این یادگیری می‌توان از روش پس‌انتشار بدون هیچ پیش‌بینی اضافی استفاده کرد.	در این یادگیری، یک پیش‌بینی‌کننده اضافی در وضعیت فعلی برای پس‌انتشار مورد نیاز است.
۴	شبکه‌ها در این یادگیری پیچیده‌تر هستند و می‌توان آن‌ها را به عنوان گروه شبکه‌ها در نظر گرفت.	در این یادگیری شبکه‌های عصبی پیچیدگی کمتری دارند یا می‌توان گفت شبکه‌های تحت این یادگیری شبکه‌های خطی ساده هستند.

ماشین است. همچنین می‌توان یادگیری متضاد را به عنوان یک الگوریتم دسته‌بندی در نظر گرفت که در آن داده‌ها را بر اساس تشابه و عدم تشابه دسته‌بندی می‌شوند.

نمونه‌های مختلفی برای این نوع رویکرد وجود دارد که یکی از نمونه‌های اصلی، چارچوب SimCLR [۷] توسط تیم هوش مصنوعی «گوگل» است. این چارچوب ابتدا نمایش‌های عمومی تصاویر را روی یک مجموعه داده بدون برچسب می‌آموزد و سپس با مجموعه داده‌های کوچکی از تصاویر برچسب‌گذاری شده برای دسته‌بندی مشخص تنظیم می‌شود. نمایش‌های اصلی با به حداکثر رساندن توافق بین نسخه‌ها یا نماهای مختلف یک تصویر و کاهش تفاوت با استفاده از یادگیری متضاد، آموخته می‌شوند. به روز رسانی پارامترهای یک شبکه عصبی با استفاده از این هدف، باعث می‌شود نمایش نماهای متناظر یکدیگر را جذب کنند و نمایش‌های نماهای غیر متناظر یکدیگر را دفع کنند. تفاوت‌های بین دو رویکرد متضاد و غیرمتضاد به طور خلاصه در جدول ۱.۴ بیان شده است.

همانطور که ذکر شد، هدف اصلی یادگیری خودنظارتی، یادگیری از داده‌های با کیفیت پایین‌تر و هدف یادگیری متضاد تمایز بین داده‌های مشابه و داده‌های غیرمشابه است. همچنین، اگر بحث کلاس‌های یادگیری خودنظارتی باشد، می‌توان دید که یادگیری خودنظارتی متضاد، کلاسی از یادگیری خودنظارتی است. در بیشتر موارد، مشاهده می‌شود که بازنمایش حاصل از یادگیری خودنظارتی برای الگوریتم‌های وظایف پایین‌دستی (مانند تشخیص چهره و تشخیص اشیا در حوزه بینایی) اعمال می‌شود. بازنمایی حاصل با عملکرد وظایف پایین‌دستی ارزیابی می‌شود. در طول این فرآیند، اطلاعات مهم و مفیدی در مورد بازنمایی آموخته شده به دست می‌آید، اما هیچ بازخوردی در

مورد این که چرا چنین عملکردی را در وظایف پایین‌دستی دریافت می‌شود، داده نمی‌شود. با استفاده از یادگیری خودنظارتی متضاد، می‌توان شهود و حدس‌هایی را برای کارایی بازنمایی آموخته‌شده به دست آورد. برای نمایش بهتر فرآیند و داده‌ها، از یادگیری متضاد با یادگیری خودنظارتی استفاده می‌شود. برای درک بازنمایی، لازم است برخی از مفاهیم بنیادی بازنمایی ذکر شود.

- اندازه‌گیری ناوردایی‌ها:

  ناوردایی در دسته‌های داده‌ها یک جزء حیاتی از بازنمایی است. این مؤلفه برای نمایش در کارهای پایین دستی بسیار مفید است.

به طور مثال برای بازنماییِ خوب داده‌های بصری، نمایش باید عمدتاً نسبت به همه تبدیل‌ها ثابت باشد. از نظر ریاضی، اگر تابع نمایش $h(x)$ باشد، باید نسبت به تبدیل t ثابت باشد:

$$t : x \to x \text{ if } h(t(x)) = h(x)$$

- داده‌افزایی و تقویت:

  یادگیری متضاد با نمونه‌های مثبت و منفی کار می‌کند و تمرکز روی یافتن نمونه‌های مثبت از داده‌ها است تا بتوان آن را به الگوریتم وظیفه‌محور پایین‌دستی ارائه داد. بیشتر اوقات شبکه برای یادگیری متضاد در زمان آموزش، از نتیجه حاصل از داده‌افزایی داده‌های آموزشی استفاده می‌کند. به عنوان مثال، در مورد حوزه بینایی کامپیوتر، بخشی از تصاویر به صورت تصادفی برش داده شده و به عنوان جفت مثبت استفاده می‌شود که یک روش ضروری برای تطبیق ویژگی‌های تصاویری است که به طور کامل قابل مشاهده نیست. این فرآیند مسئول ارائه یک نتیجه با کیفیت بالا از یادگیری خودنظارتی متضاد با داده‌افزایی است.

- بایاس دادگان:

  در یادگیری ماشین با استفاده از هر نوع یادگیری، آموزش مدل با مجموعه آموزشی انجام می‌شود. در اینجا نیز رویکردهای خودنظارتی متضاد بر روی دادگان‌های مختلف آموزش داده می‌شوند و اثراتی که در آموزش مشاهده می‌شود، می‌تواند توسط بایاس ایجاد شود. تأثیرات می‌تواند مثبت یا منفی باشد و بر نمایش داده‌ها تأثیر بگذارد. به طور مثال در حوزه بینایی، بیشتر رویکردهای خودنظارتی متضاد بر روی یک دادگان خاص، به نام ImageNet ، آموزش می‌بینند که تصاویر موجود در این دادگان، شی محور هستند. در اینجا، بازنمایی‌هایی که بایاس‌ها را از هم متمایز نمی‌کنند، می‌توانند به عملکردهای بهبود یافته دست یابند.

تا اینجا، شهود اساسی پشت یادگیری خودنظارتی متضاد و نحوه تأثیر بازنمایی و مؤلفه آن بر فرآیند مشاهده شد. حال با این توضیحات، چند نمونه از چارچوب‌های که بستری برای یادگیری خودنظارتی متضاد فراهم می‌کنند، در زیر فهرست شده است:

- چارچوب MoCO یا تضاد تکانه، چارچوبی برای یادگیری بازنمایی بصری است.

- در چارچوب PIRL ، هدف ایجاد نمایش تصویری معنادار است و این چارچوب نیازی به نمونه‌های آموزشی بالایی از تصاویر ندارد.
- چارچوب SimCLR از «گوگل»، در زمینه یادگیری خودنظارتی، نیمه‌نظارتی و دسته‌بندی تصاویر، پیشرفت‌هایی را ارائه داده است.
- چارچوب Wav2Vec ، مدلی برای پیش‌آموزش بی نظارت برای تشخیص گفتار با یادگیری بازنمایی صدای خام است که توسط تحقیقات هوش مصنوعی «فیس‌بوک» توسعه یافته است.

۲.۳.۴ روش مد نظر در این پروژه

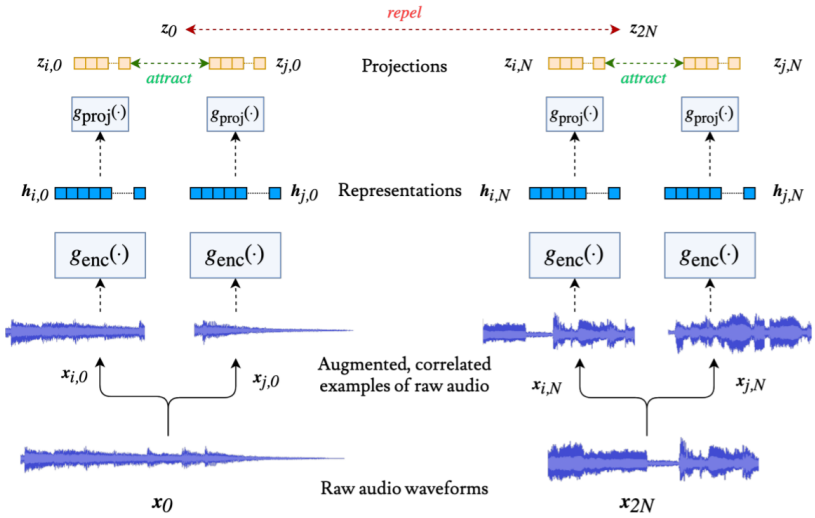
همانطور که گفته شد، یادگیری خودنظارتی موفقیت زیادی در پیشرفت حوزه‌های دیگر داشته است. اما کار بر روی یادگیری خودنظارتی در حوزه صدا و موسیقی هنوز بسیار محدود است. در این پروژه، به کمک SimCLR، که برای ساخت یک چارچوب یادگیری متضاد ساده از بازنمایی‌های بصری استفاده می‌شود، زنجیره بزرگی از داده افزایی‌های صوتی را برای پایه‌گذاری یک بستر پایه برای یادگیری متضاد بازنمایی‌های موسیقی به روش خودنظارتی، با نام CLMR [۲۲]، ارائه شده است. این رویکرد روی داده‌های موسیقی خام دامنه زمانی کار می‌کند و برای یادگیری بازنمایی‌های مفید نیازی به برچسب‌گذاری ندارد.

در واقع در این پروژه تلاش شده تا موفقیت یادگیری متضاد در یادگیری بازنمایی‌های شنیداری کارآمد نشان داده شود.

روش شناسی

هدف در SimCLR ، به حداکثر رساندن توافق بازنمایش‌های نهفته نماهای افزایش یافته یک تصویر با استفاده از کاهش تضاد است. در CLMR ، این چارچوب با حوزه صدای موسیقی خام تطبیق داده شده است. در حالی که اکثر اجزای اصلی CLMR در کارهای قبلی ظاهر شده‌است، توانایی آن در مدل‌سازی شکل‌های موج موسیقی را نمی‌توان با یک انتخاب طراحی نشان داد؛ بلکه با ترکیب آن‌ها باید توضیح داد. مدل‌های خودنظارت بر روی وظیفه برچسب‌گذاری موسیقی پایین‌دستی، ارزیابی می‌شوند و قابلیت تطبیق‌پذیری آنها ارزیابی می‌شود. برچسب‌های موسیقی بسیاری از ویژگی‌های موسیقی را توصیف می‌کنند؛ به‌عنوان مثال در دادگان نوا، نوع ساز، نوازنده و دستگاه مشخص است[۱]. ابتدا چهار مؤلفه اصلی را در بخش‌های فرعی زیر توضیح داده شده است:

- یک ترکیب تصادفی از افزایش داده‌ها که دو نمونه همبسته و افزایش یافته از قطعه صوتی یکسان، «جفت مثبت» را تولید می‌کند که با x_i و x_j نشان داده می شود.



شکل ۱.۴: چارچوب کاملی که بر روی صدای خام کار می‌کند[۲۲].

- یک شبکه عصبی رمزگذار $genc()$ که نمونه‌های افزایش یافته را به بازنمایش‌های نهفته آنها نگاشت می‌کند.
- یک شبکه عصبی پروژکتور $gproj()$ که بازنمایش‌های کدگذاری شده را به فضای پنهانی که در آن زیان متضاد فرموله می‌شود، نگاشت می‌کند.
- یک تابع زیان متضاد، که هدف آن شناسایی x_j از مثال‌های منفی در دسته $x_{k\neq i}$ برای یک x_i معین است.

چارچوب کامل در شکل ۱.۴ نشان داده شده است.

داده افزایی

همچنین، یک زنجیره جامع از داده‌افزایی صوتی برای شکل‌های موج صوتی خام موسیقی طراحی شده است تا شناسایی جفت نمونه‌های صحیح را برای مدل سخت‌تر کند. هر داده افزایی متوالی به طور تصادفی روی x_i و x_j به طور مستقل اعمال می‌شود، یعنی، هر داده افزایی دارای احتمال مستقلی است که در صوت اعمال شود. ترتیب داده افزایی‌های اعمال شده روی صدا به دقت در نظر گرفته می‌شود؛ به عنوان مثال، اعمال یک اثر تاخیری پس از طنین به طور تجربی نتیجه کاملا متفاوتی در موسیقی می‌دهد.

بدین منظور، یک قطعه تصادفی به اندازه‌ای مشخص، بدون اینکه سکوت از آن حذف شود، از یک قطعه موسیقی انتخاب می‌شود، (سکوتی که گاهی در ابتدا یا انتهای قطعه موسیقی وجود دارد) دو مثال از یک قطعه صوتی می‌توانند همپوشانی داشته باشند یا بسیار جدا و متفاوت باشند

و به مدل اجازه می‌دهد همه‌ی ساختارها را استنتاج کند. همچنین، قطبیت سیگنال صوتی معکوس شده است، یعنی دامنه در ۱- ضرب شده است. سیگنال با تاخیر ۰.۵ به سیگنال اصلی اضافه می‌شود. تأخیر به طور تصادفی بین ۲۰۰ تا ۵۰۰ میلی‌ثانیه و با افزایش ۵۰ میلی‌ثانیه نمونه‌برداری می‌شود[۲۲].

ترکیب دسته‌ای

اندازه دسته‌ای بزرگ، هدف یادگیری متضاد را سخت‌تر می‌کند؛ اما می تواند عملکرد مدل را به طور قابل ملاحظه ای بهبود بخشد[۷]. یک قطعه موسیقی از دسته برداشته می‌شود، سپس با داده افزایی، تبدیل به دو نمونه می‌شود و با آن‌ها به عنوان جفت مثبت رفتار خواهد شد. با (N-۱) ۲ نمونه باقی مانده در دسته به عنوان مثال‌های منفی رفتار شده است و نمونه‌های منفی به صراحت نمونه برداری نشده است. اندازه‌های دسته‌ای بزرگ‌تر یک مشکل عملی برای صدای خام در هنگام آموزش بر روی یک پردازنده گرافیکی ایجاد می‌کند، زیرا ابعاد ورودی آن‌ها برای نرخ نمونه بالاتر افزایش می‌یابد. هنگام آموزش بر روی چندین پردازنده گرافیکی، از نرمال‌سازی دسته‌ای سراسری استفاده می‌شود[۲۲].

رمزگذاری

در این روش، برای مقایسه مستقیم یک مدل با نظارت پیشرفته که در دسته‌بندی موسیقی بر روی شکل‌های موج خام استفاده می‌شود در مقابل یک مدل خودنظارتی، از معماری SampleCNN به عنوان رمزگذار استفاده شده است[۱۶]. به طور مشابه، از ورودی صوتی ثابت ۵۹۰۴۹ نمونه با فرکانس نمونه‌برداری ۲۲۰۵۰ هرتز استفاده شده است. بردارهای ویژگی بدست آمده از رمزگذار را می‌توان مستقیماً در یادگیری استفاده کرد، اما فرمول‌بندی هدف روی رمزگذاری‌هایی که با یک تابع پارامتری به فضای پنهان متفاوت نگاشت شده‌اند به اثربخشی نمایش‌ها کمک می‌کند[۲۲].

تابع زیان متضاد

تابع زیان متضادی که در این مدل استفاده می‌شود، زیان متقابل آنتروپی با مقیاس درجه حرارت نرمال شده است که معمولاً به عنوان زیان NT-Xent نشان داده می‌شود.

$$l_{i,j} = -log\frac{exp(sim(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} 1_{[k\neq i]} exp(sim(z_i, z_j)/\tau)}$$

شباهت زوجی با استفاده از شباهت کسینوس اندازه‌گیری می‌شود و پارامتر دما τ به مدل کمک می‌کند تا از منفی‌های سخت یاد بگیرد. تابع نشانگر $1_{[k\neq i]}$ تنها در صورتی که $k \neq i$ باشد، ۱ خواهد بود و در غیر این صورت، مقدارش ۰ خواهد بود. زیان برای همه جفت‌های (z_i, z_j) و (z_j, z_i) برای $i \neq j$ محاسبه می‌شود[۲۲].

ارزیابی خطی

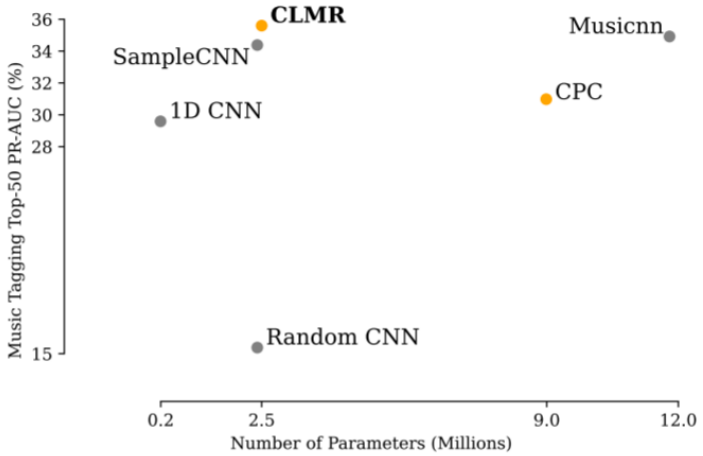
ارزیابی بازنمایی‌های آموخته‌شده توسط مدل‌های خودنظارتی معمولا با ارزیابی خطص انجام می‌شود، که در این مرحله این که کلاس‌های مربوطه تا چه حد تحت بازنمایی‌های آموخته شده به صورت خطی قابل تفکیک هستند، اندازه‌گیری می‌شود. در اینجا، بازنمایی‌هایی برای تمام نقاط داده از یک شبکه ثابت CLMR پس از همگرا شدن پیش‌آموزش به دست می‌آید و یک دسته‌بندی کننده خطی با استفاده از این نمایش‌های خودنظارت شده، در وظیفه پایین‌دستی آموزش داده می‌شود[۲۲].

بهینه سازها

در [۲۲] از بهینه ساز آدام [۱۴] با نرخ یادگیری ۰.۰۰۰۳ ، $\beta_1 = ۰.۹$ و $\beta_2 = ۰.۹۹۹$ در طول آموزش استفاده شده و از مقدار اولیه He برای همه لایه‌های هم‌گشت استفاده شده است. پارامتر دما τ روی ۰.۵ تنظیم شده است، زیرا نتایج ثابتی را بدون توجه به اندازه‌های مختلف دسته و دمای $\tau \in \{0.10.51.0\}$ مشاهده شده است. الگوریتم آدام یک الگوریتم بهینه‌سازی جایگزین برای کاهش گرادیان تصادفی برای آموزش مدل‌های یادگیری عمیق بوده و یک روش نرخ یادگیری تطبیقی است، به این معنی که نرخ یادگیری فردی را برای پارامترهای مختلف محاسبه می‌کند.

مقایسه عملکرد

همانطور که از شکل ۲.۴ آشکار است، این چارچوب، CLMR ، علیرغم پیش‌آموزش و آموزش دسته‌بندی‌کننده خطی در انجام وظیفه پایین‌دستی با سیگنال‌های خام صدا، عملکرد قوی‌تری در کار دسته‌بندی موسیقی در مقایسه با مدل‌های با نظارت، به دست می‌آورد. همچنین، CLMR دسته‌بندی کارآمد را امکان‌پذیر می‌کند؛ یعنی به کمک آن و با استفاده از بخش بسیار کمی از داده‌های برچسب‌گذاری شده، به عملکرد قابل مقایسه‌ای می‌توان رسید (عملکرد قابل مقایسه‌ای با استفاده از کمتر از یک صدم از داده‌های برچسب‌دار بدست می‌آید). CLMR می‌تواند بدون نیاز به تغییر و تنظیم دقیق در داده‌های ورودی، از هر مجموعه داده‌ای از صدای موسیقی خام بیاموزد و مدل‌ها به برچسب‌های دستی برای پیش‌آموزش نیاز ندارند. به علاوه، قابلیت انتقال خارج از دامنه بازنمایی‌های آموخته شده از قبل از آموزش CLMR را بر روی مجموعه های کاملا متفاوت صوتی موسیقی نشان داده می‌شود[۲۲].



شکل ۲.۴: مقایسه عملکرد و پیچیدگی مدل‌های با نظارت و مدل‌های خودنظارتی در طبقه‌بندی موسیقی شکل‌های موج صوتی خام روی مجموعه داده MagnaTagATune برای ارزیابی بازنمایی‌های موسیقی[۲۲].

فصل ۵

جمع بندی

به طور کلی، یادگیری با نظارت طی یک وظیفه خاص با یک دادگان بزرگ با برچسب‌گذاری دستی که به طور تصادفی به مجموعه آموزش، اعتبارسنجی و مجموعه آزمایش تقسیم می‌شود، آموزش داده می‌شود. با این حال، یادگیری با نظارت نه تنها به شدت به برچسب‌گذاری دستی پرهزینه متکی است، بلکه از خطای تعمیم، روابط همبستگی نادرست و حملات خصمانه نیز رنج می‌برد. انتظار می‌رود که شبکه عصبی با برچسب‌ها، نمونه‌ها یا آزمایش‌های کمتر، بیشتر و بهتر بیاموزد. به عنوان یک نامزد امیدوارکننده، یادگیری خودنظارتی توجه زیادی را به دلیل کارایی بهتر داده و توانایی تعمیم فوق‌العاده‌اش به خود جلب کرده است و بسیاری از مدل‌های پیشرفته از این الگو پیروی می‌کنند.

در این روش، تلاش بر این بود که نشان داده شود که نمایش‌های CLMR با استفاده از مجموعه داده‌های خارج از دامنه قابل انتقال هستند. این امر نشان می‌دهد این روش از تعمیم‌پذیری بالایی در طبقه‌بندی موسیقی برخوردار است. همچنین، هدف این بود که نشان داده شود که روش پیشنهادی برای یادگیری کارآمد داده دادگان‌های برچسب‌دار کوچک‌تر قابل اعمال است و بدین منظور سعی شد از دادگان نوا که شامل تعداد زیادی قطعات تکنوازی از پنج ساز موسیقی سنتی ایرانی و برچسب‌دار هست، برای آموزش و تعمیم به دادگان جدید استفاده شده است.

گام‌های اولیه نیز برای ارائه یک مدل پیش‌آموزش داده شده برداشته شد و از رویکرد خودنظارتی متضاد بدین منظور استفاده و بررسی شد. در این پژوهش سعی بر این بود تا ثابت شود که به کارگیری همزمان یادگیری خودنظارتی و یادگیری متضاد در حین آموزش، بازنمایی‌های آموخته شده را بهبود می‌بخشد.

دغدغه اصلی در این حوزه، نبود دادگان با جامعیت و تنوع کافی بود؛ که در این پژوهش سعی شده، با جمع آوری مجموعه‌ای جامع و با تنوع بالا و به کمک یادگیری خودنظارتی و با پایه گذاری یک بستر پایه، در راستای تشخیص ساز و آواز سنتی ایرانی، قدم برداشته شود. دادگان این پژوهش به منظور تشخیص ساز، جمع آوری و طراحی شد و در طراحی آن سعی شد که از حجم و تنوع کافی

به لحاظ نوع ساز، هنرمند و دستگاه برخوردار باشد. همه قطعات موسیقی موجود در این دادگان، تک‌کانال، به فرم $mp3$ ، با فرکانس نمونه‌برداری ۲۲۰۵۰ هرتز و تفکیک‌پذیری بیتی ۱۶ بیت است. دادگان گردآوری شده شامل ۱۶۸۹۵ قطعه از ۱۸۱ هنرمند و در حدود ۹۶۰ ساعت است و شامل قطعات موسیقی سنتی و تلفیقی سنتی ایرانی است. با توجه به هدف این پژوهش، که تشخیص ساز در قطعه موسیقی است، نیاز به بررسی فراوانی استفاده از سازهای مختلف در موسیقی سنتی بود، تا سازهای پرتکرار به طور حدودی شناسایی شود؛ بدین منظور ۱۰۰ آلبوم به صورت تصادفی در نظر گرفته شد و با توجه به اطلاعات منتشر شده رسمی در اطلاعات آلبوم‌ها، بررسی شد. در ۱۰۰ آلبوم بررسی شده اولیه، توزیع فراوانی سازهای استفاده شده به ترتیب به شرح زیر است:

جدول ۱.۵: توزیع فراوانی سازهای استفاده شده در بخشی از دادگان

تنبک	سنتور	کمانچه	نی	تار	سه تار	عود	دف	قیچک	دایره
۶۴	۴۶	۴۵	۳۹	۳۷	۳۴	۳۲	۳۰	۱۴	۱۲

رباب	کوزه	قانون	صراحی	دو تار	دهل	کمان	سبو	کاسه
۱۰	۴	۳	۳	۲	۲	۲	۱	۱

قوشمه	طبل	دمام	شمشال	کاخن	سرنا	تنبور	پرهیب	دیوان
۱	۱	۱	۱	۱	۱	۱	۱	۱

همانطور که از نتایج جدول مشخص است، مرز بین سازهای پرتکرار کاملا مشخص بوده و می‌توان بدون کاهش کلیات برچسب، سازهای پرتکرار را یازده ساز اول در نظر گرفت و از سایر سازها مانند سازهای موسیقی کلاسیک غربی که در موسیقی تلفیقی به کار رفته‌اند؛ مثل پیانو، ویولن، ویولن سل، می‌توان صرف نظر کرد. در ادامه، جدولی از اسامی هنرمندان صاحب‌اثرِ قطعات موجود در دادگان نیز ارائه شده است(جدول ۲.۵).

پیشنهادی برای پژوهش‌های آینده این است که نشان داده شود که استفاده همزمان از یادگیری با نظارت و متضاد در حین آموزش، علاوه بر این که بازنمایی‌های آموخته شده را بهبود می‌بخشد، به آموزش نیز سرعت می‌بخشد. همچنین، می‌توان تلاش کرد تا در استفاده از روش‌های یادگیری خودنظارتی، از دادگانی با حجمی کمتر استفاده و نتایج را مقایسه کرد. هدف این است که ماشین به منظور یادگیری، با دادگانی کم حجم و بدون برچسب‌گذاری و حاشیه‌نویسی با دخالت انسان، به پردازش و بازیابی اطلاعات بپردازد.

پیشنهاد دیگر نیز این است که به جای استفاده از یادگیری خودنظارتی متضاد، ایده‌های یادگیری خودنظارتی و دیگر روش‌های یادگیری، مانند یادگیری چند وظیفه‌ای، برای پیش‌آموزش، ترکیب شوند. در این رویکرد، ورودی‌های صوتی خام از چندین لایه رمزگذاری عبور می‌کنند و خروجی‌ها نمایش‌های دو بعدی با اطلاعات زمانی هستند.

جدول ۲.۵: جدول هنرمندان صاحب‌اثر قطعات دادگان

همایون شجریان	کیوان ساکت	ایرج بسطامی	حشمت رجب زاده	حسین حمیدی
وحید تاج	هومن مهدویان	حمید قنبری	سعید نایب محمدی	احسان ذبیحی فر
محمد معتمدی	مجتبی عسگری	پیام جهانمانی	صهبا مطلبی	حسین علیشاپور
سیامک جهانگیری	فرزاد فضلی	نگین زمردی	امیر شریفی	علی کاظمی
علی اکبر مرادی	محمدرضا ابراهیمی	نگار خارکن	علیرضا فلسفی	هومان رومی
علی مومنیا	گروه دستان	سالار عقیلی	سام امیرحسین	مسعود شعاری
نورعلی برومند	عیسی غفاری	حسن خان	هوشنگ کامکار	پرویز مشکاتیان
علی جهاندار (گروه عارف)	فرامرز پایور	حسین علیزاده	شهرام ناظری	پرواز همای
جلال ذوالفنون	محمد موسوی	سیدعبدالحسین مختاباد	جلال تاج اصفهانی	علیرضا افتخاری
علیرضا قربانی	اردشیر کامکار	محمدرضا لطفی	غلامحسین بنان	روح الله خالقی
گروه موسیقی هفت خوان	فرید خردمند	حسام اینانلو	پیمان یزدانیان	همایون نصیری
سینا علم	سحاب علم	سیامک ایقانی	پدارم بلوچی	شهرام میرجلالی
لطف الله مجد	علی پژوهشگر	بهروز همتی	رضا پرویززاده	سعید نایب محمدی
ایرج رحمان پور	علی اکبر شکارچی	حسین پرنیا	فاضل جمشیدی	حسین رضا اسدی
حسین تهرانی	مهرداد پیکرزاده	مرتضی محجوبی	اسدالله ملک	ابوالحسن صبا
رضا ورزنده	علی اصغر بهاری	حسن کسائی	جلیل شهناز	علی رستیمان
سعید خلج	مجید درخشانی	امیرعباس ستایشگر	سیاوش ایمانی	حسن خدایی نیا
سهیل حکمت آرا	فردین خلعتبری	فرشاد جمالی	تهمورس پورناظری	بهرام حصیری
حسام الدین سراج	رضا قلی میرزا ظلی	پژمان حدادی	بابک شریفی مجد	پویا سرایی
مصباح قمصری	پویان بیگلر	پریچهر خواجه	بهنام معصومی	سامر حبیبی
مجید خلج	حسین عمومی	فرهاد فخرالدینی	صدیق تعریف	بهرام با جلان
عطا جنگوک	بهرام سارنگ	ناصر فرهنگ فر	داریوش طلایی	محمدرضا درویشی
حمیدرضا نوربخش	داریوش پیرنیاکان	گروه کامکارها	بیژن کامکار	هوشنگ کامکار
سیدعلی‌اصغرکردستانی	علی رهبری	علی قمصری	انوشیروان روحانی	سلمان سالک
مجید علیزاده	علیرضا خشتی	زکریا یوسفی	حمید متبسم	رسول اکبری
همایون خرم	سامان احتشامی	سیامک بنایی	علیرضا گلبانگ	علیرضا حاجی طالب
محسن کرامتی	محمدباقر زینالی	محمدامین اکبرپور	رامین بحیرائی	سجاد پورقناد
رضا صبری	حسین بهروزی نیا	اشکان کمانگری	جاوید افسری راد	داوود آزاد
علی انصاری	صائب کاکاوند	اردوان کامکار	مسعود جاهد	ارسلان کامکار
سیاوش کامکار	فردین کریم خاوری	آرش قاسمی	گروه چارتار	علی زند وکیلی
حسین خواجه امیری (ایرج)	آرین رحمانیان	فرشاد عباسی	بهرام دهقانیار	ابوالفضل صادقی نژاد
بامداد ملکی	علیرضا برزگر	گروه بال و شال	ابولحسن اقبال آذر	شهاب اکبری
حسین مهرانی	سپهر سراجی	گروه بوم	بهروز رونده	ایرج صهبایی
محمود حشمت	علی اصغر شاه زیدی	جمشید پورمهر	سهیل مخبری	مهیار طریحی
محمد اسماعیلی	محمد شمس	رامین بحیرائی	کیوان علی محمدی	سیامک شجریان
ابوسعید مرضایی	علیرضا گلشن	فرید یداللهی	مرتضی فلاحتی	میرزا علی چهرازی

واژه‌نامه

Ethnomusicology	اتنوموزیکولوژی
Spectral Centroid (SC)	انحراف معیار سنترویید
Echelle	اِشِل
Bias	بایاس
Batch composition	ترکیب دسته‌ای
Generalizability	تعمیم‌پذیری
channel Mono	تک‌کانال
Discriminative	تمایزی
Radial Basis Functions	توابع شعاعی پایه
Four main tone attributes	چهار رکن اصلی و بنیادی
Clustering	خوشه‌بندی
Classification	دسته‌بندی
Dynamics (Loudness, Intensity)	دینامیک (شدت صوت یا دامنه)
Mel-frequency Cepstral Coefficient (MFCC)	ضرایب کپسترال بر مبنای مقیاس مل
Timbre (Tone color)	طنین (رنگ صوتی یا شیوش)
Spectrogram	طیف‌گرام
Carnatic	کارناتیک
Duration (Note Value/Time Interval)	کشش
Support Vector Machine	ماشین بردار پشتیبان
Generative	مولد
Invariance	ناوردایی
Inharmonicity	ناهمگونی
Pitch	نواک
Downstream tasks	وظایف پایین‌دستی
Convolution	هم‌گشت

Supervised learning	یادگیری با نظارت
Unsupervised learning	یادگیری بی نظارت
Self-supervised learning	یادگیری خودنظارتی
Deep learning	یادگیری عمیق
Machine learning	یادگیری ماشین
Contrastive learning	یادگیری متضاد
Semi-supervised learning	یادگیری نیمه‌نظارتی
K Means	K میانگین
K Nearest Neighbor	K تا نزدیکترین همسایه

مراجع

[۱] باقر باباعلی، آشنا گرگان محمدی و اسماء فرجی دیزجی. «نوا: دادگان موسیقی سنتی ایرانی برای تشخیص دستگاه و سازهای اصیل ایرانی.» پردازش سیگنال پیشرفته ۳، ۲، دانشگاه تهران، ۱۳۹۸

[۲] سار محمودان، ایوب بنوشی، «دسته‌بندی خودکار گام ماهور موسیقی ایرانی توسط یک شبکه عصبی مصنوعی»، دومین کنفرانس بین المللی آکوستیک و ارتعاشات، دانشگاه صنعتی شریف، ۱۳۹۱

[۳] صابر عبدالله زادگان، شهرام جعفری، مرتضی دیرند، «تشخیص خودکار دستگاه و گام موسیقی سنتی ایرانی مبتنی بر تکنوازی سازهای تار و سنتور به وسیله استخراج نت هوشمند»، بیستمین کنفرانس ملی سالانه انجمن کامپیوتر ایران، دانشگاه فردوسی مشهد، ۱۳۹۳

[۴] Abdoli, Sajjad. "Iranian Traditional Music Dastgah Classification." ISMIR. 2011.

[۵] Beigzadeh, Borhan, and Mojtaba Belali Koochesfahani. "Classification of Iranian traditional musical modes (DASTGÄH) with artificial neural network." Journal of Theoretical and Applied Vibration and Acoustics 2.2 (2016): 107-118.

[۶] Chapelle Olivier, Schölkopf Bernhard, Zien Alexander. "Semi-Supervised Learning", Cambridge, MIT Press Scholarship Online, 2006.

[۷] Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PMLR, 2020.

[۸] Darabi, Nima, N. Azimi, and Hassan Nojumi. "Recognition of Dastgah and Maqam for Persian music with detecting skeletal melodic models."

The second annual IEEE BENELUX/DSP Valley Signal Processing Symposium. 2006.

[۹] Engel, Jesse, et al. “Neural audio synthesis of musical notes with wavenet autoencoders.” International Conference on Machine Learning. PMLR, 2017.

[۱۰] Ericsson, Linus, et al. “Self-Supervised Representation Learning: Introduction, Advances and Challenges.” arXiv preprint arXiv:2110.09327 (2021).

[۱۱] Gfeller, Beat, et al. “Pitch estimation via self-supervision.” ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020.

[۱۲] Hajimolahoseini, Habib, Rassoul Amirfattahi, and Maryam Zekri. “Real-time classification of Persian musical Dastgahs using artificial neural network.” The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012). IEEE, 2012.

[۱۳] Heydarian, Peyman. “Automatic recognition of Persian musical modes in audio musical signals”. Diss. London Metropolitan University, 2016.

[۱۴] Kingma, D., and J. Ba. ”Dp kingma and j. ba, adam: A method for stochastic optimization.” arXiv preprint arxiv:1412.6980 (2015).

[۱۵] LAYEGH, Mahmood ABBASI, Siamak HAGHIPOUR, and Yazdan 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 1.4 (2013): 57-66.

[۱۶] Lee, Jongpil, et al. “SampleCNN: End-to-end deep convolutional neural networks using very small filters for music classification.” Applied Sciences 8.1 (2018): 150.

[۱۷] Pascual, Santiago, et al. “Learning problem-agnostic speech representations from multiple self-supervised tasks.” arXiv preprint arXiv:1904.03416 (2019).

[۱۸] Ravanelli, Mirco, et al. “Multi-task self-supervised learning for robust speech recognition.” ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020.

[۱۹] RezezadehAzar, Shahla, et al. “Instrument-Independent Dastgah Recognition of Iranian Classical Music Using AzarNet.” arXiv preprint arXiv:1812.07017 (2018).

[۲۰] Saeed, Aaqib, David Grangier, and Neil Zeghidour. “Contrastive learning of general-purpose audio representations.” ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021.

[۲۱] Sanati, Farshad. “An investigation on the value of intervals in Persian music.” (2020).

[۲۲] Spijkervet, Janne, and John Ashley Burgoyne. “Contrastive learning of musical representations.” arXiv preprint arXiv:2103.09410 (2021).

[۲۳] Tagliasacchi, Marco, et al. “Pre-training audio representations with self-supervision.” IEEE Signal Processing Letters 27 (2020): 600-604.

[۲۴] Talai, Dariush. “Radif analysis: based on the Notation of Mirza Abdollah’s Radif with Annotated visual description”. Tehran, Ney Publishing Co. 2015.

[۲۵] Vafaeian, Amir, et al. “Automatic Identification and Classification of the Iranian Traditional Music Scales (Dastgāh) and Melody Models (Gusheh): Analytical and Comparative Review on Conducted Research.” Human Information Interaction 5.2 (2018): 46-72.

Abstract

Automatic identification and classification of Dastgahs, instruments and Gushehs of Iranian traditional music is one of the branches of music information retrieval. There is a lot of research in the field of music signal processing in the world in order to retrieve information from music based on content. Unfortunately, there is very little research on computer processing of Iranian traditional music. Since most musical instrument recognition databases focus on Western musical instruments, it is difficult for researchers to study and evaluate the field of Iranian traditional musical instrument recognition. Most of the research that has been published so far, based solely on eschelle of the five main Dastgahs, has been done in order to automatically distinguish and identify these Dastgahs from each other. Since the classification of Dastgahs does not have the necessary originality, and in this regard there is no consensus among theorists and musicians in terms of the number of instruments and the boundaries between them, more research and with new methods is still needed in this field. One of the main reasons for the lack of research in the field of traditional Iranian music can be considered the lack of data. This research work has been done with the aim of gathering a comprehensive and diverse database for a basic issue in the field of Iranian traditional music.

College of Science
School of Mathematics, Statistics, and Computer Science

Recognition of the Type and Number of Instruments in Iranian Traditional Music

Sahar Sadat Shirmardi

Supervisor: Dr. Bagher Babaali

A thesis submitted in partial fulfillment of the requirements for
the degree of B.Sc. in Computer Science

Februray 2022