ششمین کنفرانس بین‌المللی
«بازی‌های رایانه‌ای؛ فرصت‌ها و چالش‌ها»
۳۰ بهمن و ۱ اسفند ۱۳۹۹ – دانشگاه اصفهان

بررسی امکان تولید خلاقانه‌ی موسیقی متن بازی‌های رایانه‌ای با استفاده از هوش مصنوعی مبتنی بر مدل مخلوط گاوسی

بهنام علیزاده اشرفی[۱]، کاظم پورالوار[۲]، مرضیه سادات مطلقی[۳]*

۱- **استادیار و عضو هیئت علمی دانشکده چند رسانه ای، دانشگاه هنر اسلامی تبریز**
b.alizadehashrafi@tabriziau.ac.ir

۲- **استادیار و عضو هیئت علمی دانشکده چند رسانه ای، دانشگاه هنر اسلامی تبریز**
Email: k.pouralvar@tabriziau.ac.ir

۳- **دانشجوی کارشناسی ارشد هنرهای رایانه ای، دانشگاه هنر اسلامی تبریز**
Email: m.motlaghi@tabriziau.ac.ir

چکیده

هنر و در راس آن موسیقی، وسیله‌ای برای ابراز درونیات انسان است که از روحش تراوش می‌کند. هنر بخشی از وجود انسان است که حاصل آن می‌تواند از تمام ابعاد درونی انسان باشد؛ می‌تواند نتیجه‌ی یک احساس و یک فکر را با هم درآمیزد و موجودیت جدیدی خلق کند. اما مشکل این‌جاست که ذهن و فکر انسان محدود است و نمی‌تواند به‌هنگام خلق یک اثر هنری، تمامی حالات و شرایط را درنظر بگیرد. پیشرفت تکنولوژی و کامپوترهای شخصی به هنرمندان کمک شایانی کرده است تا طرز فکر و درونیات خود را توسط الگوریتم‌های کامپیوتری بازتولید کنند و اثر جدیدی خلق کنند. الگوریتم‌های هوش مصنوعی قدرت بسیار زیادی به ما می‌دهند تا بتوانیم تصورات خود را با مقیاس بسیار بزرگ از نظر حجم پردازشی تولید کنیم. امروزه با گسترش کاربردهای هوش مصنوعی و فیزیک دیجیتال در عرصه‌ی تولید محتوای خلاق، درهای جدیدی برای طراحان بازی‌های دیجیتال گشوده شده است. بهره‌گیری از این فناوری در ساخت بازی‌های جدی، برای ارتقای سطح کیفی و رفع محدودیت‌های موجود، بسیار موثر واقع شده است.

مقاله‌ی حاضر با انجام مطالعاتی به بررسی امکان تولید خودکار موسیقی خلاقانه مطابق با درجه سختی بازی‌های رایانه ای با استفاده از الگوریتم گاوسی در هوش مصنوعی می‌پردازد.

کلمات کلیدی: موسیقی، هوش مصنوعی، تولید خودکار، مدل مخلوط گاوسی

۱–مقدمه

هوش مصنوعی (AI[۱]) که برخی مواقع هوش ماشینی نامیده می‌شود، به هوشمندی نشان داده شده توسط ماشین‌ها در شرایط مختلف اطلاق می‌شود که در مقابل هوش طبیعی انسان‌ها قرار دارد. به عبارت دیگر هوش مصنوعی به سامانه‌هایی گفته می‌شود که می‌توانند واکنش‌هایی مشابه رفتارهای هوشمند انسانی از جمله درک شرایط پیچیده، شبیه‌سازی فرآیندهای تفکری

[1] Artificial Intelligence

و شیوه‌های استدلالی انسانی و پاسخ موفق به آنها، یادگیری و توانایی کسب دانش و استدلال برای حل مسائل را داشته باشند [1].

امروزه، هوش مصنوعی به عنوان یکی از بالقوه ترین و تأثیرگذارترین حوزه‌های فناوری در حیطه‌ی موسیقی، به ویژه برای آهنگسازی شناخته شده است. هوش مصنوعی به عنوان روشی نوین برای تقویت خلاقیت انسان، توجه گسترده متخصصان و علاقه‌مندان را جلب می کند. برخی آهنگسازان معتقدند فناوری‌های نوظهور چندین بار زمینه موسیقی را تغییر شکل داده‌اند، درک آن‌ها از خلاقیت را به چالش کشیده و ارزش های ثابت شده را زیر سوال برده‌اند [2].

امروزه موسیقی و هوش مصنوعی به واسطه‌ی پژوهش‌های انجام شده در هر دو حوزه به هم نزدیک‌تر شده‌اند. ساخت موسیقی با استفاده از راه‌کارهای موجود در هوش مصنوعی، زمینه پژوهشی جدیدی را ایجاد کرده است. علاوه بر آن ساخت خودکار موسیقی علاوه بر اینکه ما را در شناخت بهتر چگونگی تفکر موسیقیایی انسان یاری خواهدکرد، کمک شایان توجهی به آهنگ‌سازان و موسیقی‌دانان در بهبود تئوری‌های موسیقی با استفاده از قدرت محاسباتی کامپیوترها می‌کند و آهنگ‌سازان را نیز از آهنگ‌سازی‌های «نُت به نُت» رها خواهدساخت[1]. درواقع پیشرفت در تکنیک‌های یادگیری ماشین، منجر به توسعه سیستم‌های خلاق شده‌است که سطوح بالاتری از اتوماسیون را برای ایجاد موسیقی، تولید و مصرف ارائه می‌دهند. به طور کلی در زمینه ی تولید موسیقی، بسیاری از ابزارها در بازار موجود هستند که از یادگیری ماشین برای پردازش صوتی به کمک کامپیوتر استفاده می‌کنند [3]. با این حال، پیشرفت در هوش مصنوعی در طی دهه ی گذشته و در درجه ی اول در زمینه‌ی یادگیری عمیق، باعث افزایش چشم اندازها و کاهش موانع برخی از برنامه‌های تولید موسیقی شده است [4].

روند تولید موسیقی با استفاده از فناوری هوش مصنوعی براساس مجموعه داده های عظیمی از اسکریپت‌های موسیقی ساخته شده است که یک الگوریتم از آن‌ها یاد می‌گیرد. هرچند که هر فعالیت هنری شامل دانش پیشرفته حوزه‌ای است که فرد در آن فعالیت می‌کند، همچنین به طور کلی دارای دانش و تجربه‌ی گسترده‌ای از فرهنگ بشری است که از طریق یک اثر هنری تفسیر می‌شود [2]. ظرفیت خلاقانه‌ی آثار هنری مبتنی بر هوش مصنوعی توسط Manovich به خوبی توضیح داده شده است: "یک کامپیوتر باید تقریباً همان دانش یک انسان بزرگسال از جهان را داشته باشد"[2]. تکنیک‌های یادگیری ماشینی به طور مداوم زمینه‌های هنری را تحت‌تاثیر قرار می‌دهند که به طور سنتی حوزه ی افکار و عواطف انسانی هستند و پتانسیل بالایی در افزایش کارایی تولید رسانه‌های هنری نشان داده‌اند .

یکی از کاربردهای هوش مصنوعی در زمینه‌ی ساخت بازی‌های دیجیتال برای ایجاد حس غوطه‌وری و لذت بیشتر به همراه هیجان مخاطب در بازی است. طراحی بازی رایانه‌ای، صنعتی است که به شدت به نوآوری‌های هنری مانند نقاشی، مدل سازی سه‌بعدی، ترکیب موسیقی متن و غیره متکی است. آثار هنری همیشه پرمشغله قلمداد می‌شوند و توسعه بازی‌ها اغلب نیاز به تیم‌های بسیار قوی و بزرگ دارد و بعلاوه گران است. هوش مصنوعی، به عنوان همتای فیزیکی آن مانند ماشین‌های خودکار، این توانایی را دارد که تا حد زیادی کارهای انسانی را کاهش دهد. به همین منظور با توجه به وسعت توسعه‌ی حس غوطه وری بازی و محیط های گوناگون در بازی‌های رایانه‌ای به خلاقیت در ترکیب موسیقی متن نیاز داریم. از آنجا که آهنگسازان انسان‌هایی گران‌قیمت هستند، تکنیک‌های ساخت هوش مصنوعی در چندین پروژه منبع باز پیشنهاد شده‌است. فن‌آوری‌های فعلی، اجراهای خوبی را در ترکیبات ملودی کوتاه دارند اما با چالش‌های بزرگی در سطح صنعتی قطعات بسیار ساختاریافته برای بازی‌ها مواجه هستند [5]. یک بازی رایانه‌ای ممکن است روی پلان متمرکز باشد و یا اکشن باشد و یا ممکن است به سادگی در محیط بسته متمرکز باشد و در عین حال با تکنیک‌های کامپیوتری از دید بازیکن نامتناهی بنظر برسد.

موسیقی متن بازی خوب ممکن است روایی باشد، یا دارای ریتم‌های متغیر و سیناپسی باشد یا به سادگی موسیقی محیطی باشد. در یک بازی رایانه‌ای، سبک های مختلف موسیقی را می توان در مراحل مختلف با توجه به درجه‌های سختی بازی تنظیم کرد. به عنوان مثال، موسیقی متن در یک مرحله از یک بازی می تواند متفاوت از مراحل دیگر باشد.

۲– پیشینه ی پژوهش

در طول تاریخ، موسیقی بازتاب برجسته‌ای از فرهنگ بشر بوده است که با رسیدن به گوش فرد، یک ترکیب مجموعه‌ای از معیارهای اجتماعی و تکنولوژیکی عصر را ارائه می دهد که قطعه موسیقی از آن‌ها گرفته شده است. فناوری ها، به نوبه خود ارتباط مستقیمی با نحوه تولید، پخش و توزیع موسیقی دارند. از جمله اختراعاتی مانند رادیو و سنتز صدا در میان اختراعات دیگر، نیروی محرکه اصلی صنعت موسیقی بوده است. در این میان با ظهور تکنولوژی‌های جدید، به نظر می رسد هوش مصنوعی این امکان را دارد که بار دیگر صنعت را از بسیاری جهات تغییر دهد. هوش مصنوعی اصطلاحی جهانی برای الگوریتم های سازگار است، که به کارهای چالش برانگیز عقلی اختصاص داده می‌شود. در حال حاضر، توسعه در هوش مصنوعی، کاربردهای مختلفی در صنعت موسیقی پیدا کرده است؛ از الگوریتم های مورد استفاده برای توسعه موسیقی در سیستم های عامل جریان گرفته تا مواردی که قادر به ویرایش صداهای ضبط شده هستند [2]. علاوه بر این، موسیقی هوش مصنوعی صفت جدیدی از تولید فرهنگی را نشان می‌دهد: افزایش اعتماد به منابع محاسباتی مبتنی بر ابر و مجموعه داده‌های بزرگ که با تلاش‌های فردی به سختی قابل دستیابی است. در مقابل، به عنوان مثال، تخصص فردی از طریق آموزش نهادینه، تا حدودی اهمیت خود را از دست می‌دهد. مجموعه داده‌ها و نمونه‌های محاسباتی با پردازنده گرافیکی به عنوان ابزاری عمل می‌کنند که مفهوم خلاقیت را شکل می‌دهند. دسترسی به هر دو مجموعه داده و قدرت محاسباتی برای "دموکراتیزه کردن" چنین تغییری باید گسترش یابد [4].

سال‌های اخیر شاهد پیشرفت در کاربرد فناوری هوش مصنوعی مانند تکنیک‌های یادگیری عمیق برای تصمیم‌گیری خلاقانه در ساخت موسیقی، هم از نظر عملکرد بهتر و افزایش دسترسی به ابزارها و هم افزایش همزمان منافع صنعتی در انجام تحقیقات و ساخت محصولات از آن بوده‌ایم. موسیقی مبتنی بر هوش مصنوعی، راه‌اندازی فاز جدیدی از تکنولوژی دیجیتال موسیقی را نشان می‌دهد؛ چیزی که مشخص می‌کند این دوره، اتوماسیون داده‌ها در فرآیندهای ترکیب و مهندسی صوت است. با توجه به عملکرد خلاقانه، موضوعات تحقیق در زمینه‌ی موسیقی هوش مصنوعی شامل راحتی در مفهوم اتوماسیون کار انسانی، اکتشافات اشکال جدید، و ابزارهای کمکی می‌باشد .

تحقیقات در زمینه‌ی تولید موسیقی با استفاده از هوش مصنوعی، مدت‌ها قبل از آن‌که رایانه‌ها شکل آشنایی پیدا کنند، آغاز شده است. از سال ۱۹۵۰ این رشته متخصصانی را به خود جذب کرده است که الگوریتم‌ها را برای انجام کارهای مختلف موسیقی مانند آهنگسازی، بداهه‌نوازی، تشخیص موسیقی و علامت‌گذاری و بسیاری کارهای دیگر برنامه‌ریزی کرده‌اند [2] .

فرآیند تولید موسیقی هوش مصنوعی مبتنی برمجموعه‌ای از پایگاه های داده‌ای عظیمی از نت‌های موسیقی است که با استفاده از الگوریتم‌های یادگیری عمیق و روش‌های مشابه پیاده‌سازی می‌گردد. با وجود اینکه هر فعالیت هنری شامل دانش پیشرفته‌ی حوزه‌ای است که فرد در آن کار می‌کند و همچنین نیازمند دانش و تجربه گسترده از فرهنگ انسانی است که از طریق یک اثر هنری تفسیر می‌شود، با این حال ظرفیت خلاقانه آثار هنری مبتنی بر هوش مصنوعی، به‌خوبی تشریح شده‌است: "خلق آثار هنری زیبا و دلپذیر در مورد انسان‌ها و دنیایشان، ممکن است پس از پیشرفت کافی در این آثار، امکان پذیر شود".

به عبارت دیگر، یک کامپیوتر باید تقریبا همان دانش را به عنوان یک انسان بالغ داشته باشد [6]. در همین راستا بسیاری از تکنیک‌های ساخت اتوماتیک، به منظور ساخت موسیقی با استفاده از فناوری هوش مصنوعی به‌کارگرفته شده است. برای مثال، اولین برنامه آهنگسازی خودکار، مربوط به بازی تاس موتزارت در قرن هجدهم است. در این بازی، موسیقی والس با ترکیب قطعات ملودی از پیش ساخته‌شده ایجاد می‌شود [5]. گروهی دیگر از آهنگسازان خودکار آهنگ‌ها را با ساختارهای ریاضی تقلید می‌کنند، به‌طوری که انتظار می‌رود تجربه موسیقی کمی از آهنگساز انسانی حاصل شود. روش اول مبتنی بر روش‌های تحلیل الگو مانند شبکه‌های عصبی پرسپترون چند لایه mlp[1] است[7]. چه بسا امروزه آن‌ها از شبکه‌های عصبی مانند شبکه‌های عصبی بازگشتی و حافظه بلندمدت استفاده کرده‌اند [8]. تعدادی از پروژه‌های منبع‌باز هم از RNN[2] و LSTM[3] برای ساخت موسیقی استفاده کرده‌اند. در برخی از آثار، سازنده ملودی از الگوریتم‌های با نظارت و یا بدون نظارت بهره می‌برد و در اغلب مواقع الگوریتم‌های بدون نظارت توسط هنرمندان انسانی و هوش مصنوعی کنترل نمی‌شوند [9].

مطالعات نشان می‌دهد که تولید موسیقی خودکار، چالش بزرگی برای تکنیک‌های هوش مصنوعی است. صفحه موسیقی معمولاً به عنوان ابزارهایی با ساختار درختی مانند musicXML ذخیره می شود. بنابراین تعداد زیادی از روش‌های متداول یادگیری ماشین مانند شبکه‌های عصبی پیش‌خور، ماشین‌های بردار پشتیبان (svm[4]) و فرآیند گاوسی از وظایف پردازش ملودی حذف می‌شوند. شبکه‌های عصبی بازگشتی می‌توانند الگوهایی با طول متغیر را پردازش کنند، اما در کار ساخت خطوط صدا برای بازی‌ها با دو مشکل عمده مواجه هستند: اول اینکه آن‌ها سیستم‌های تابع تکرار غیرخطی blackbox (IFS) هستند که در آن مداخله‌ی انسان در ساخت ملودی کار دشواری می‌باشد. در واقع تحمیل ساختارهای جهانی به دلیل ارتباط پیچیده بین پارامترهای IFS و نمودارهای عملکردی دشوار است. دوم اینکه در هنگام ساخت آهنگ‌های طولانی، شبکه‌های عصبی بازخوردی از ابتکار خارج می‌شوند و ملودی‌ها را با تغییرات عاطفی کمی تولید می‌کنند [5].

اگرچه قابل بحث است که آیا الهامات هنری را می‌توان به درستی توسط هوش مصنوعی مدل‌سازی کرد یا نه، با این حال تکنیک‌های یادگیری مبتنی بر ماشین به طور مداوم زمینه‌های هنری را تحت‌تاثیر قرار می‌دهند و پتانسیل بالایی در افزایش کارایی تولید رسانه‌های هنری دارند. با وجود علاقه زیاد محققان در طراحی‌های هوش مصنوعی و طراحی‌های ویدیویی، مساله تولید خودکار موسیقی توجه کمتری را به خود جلب کرده‌است. به همین منظور در این مقاله برآنیم به بررسی توانایی تولید خودکار موسیقی متن خلاقانه با استفاده از فناوری هوش مصنوعی مبتنی بر یادگیری ماشین در بازی های رایانه ای بپردازیم

۳– معرفی راهکارهای هوش مصنوعی در تولید موسیقی

سیستم‌هایی که با استفاده از هوش مصنوعی به آهنگسازی می‌پردازند از راهکارهای متعددی بهره می‌گیرند. بیشتر روش‌های استفاده شده عبارتند از: مدل‌های ریاضی، سیستم‌هایی برپایه دانش، گرامرها، مدل‌های تکاملی، سیستم‌های یادگیرنده، و سیستم‌های ترکیبی [10].

در ادامه به شرح تعدادی از این روش ها و بررسی معایب آنها میپردازیم.

۱) مدل‌های مارکوف (hmms[5]):

[1] Multiplayer perceptron
[2] Recurrent Neural Network
[3] Long Short-Term Memory
[4] Support vector machine
[5] Hidden markov model

در مراحل اولیه آهنگسازی با استفاده از هوش مصنوعی، مدل‌های مارکوف (hmms) محبوب بودند؛ زیرا در مدل‌سازی و پیش‌بینی دنباله‌های زمانی موثر هستند. آهنگ‌ساز سایبرناتیک کـه به وسیله ی ایمز و دومینو در سال ۱۹۹۲ سـاخته شـد یک مثال بارز از چنین سیستمی است. این برنامـه قطعه هـای مختلف از ژانرهای مختلـف را ترکیب می‌کنـد. یکی از ویژگی‌های جالب این سیستم این است که اول ریتم ملودی را با استفاده از زنجیره‌های مارکوف استنباط می‌کند و سپس، زیروبمی‌های صـدا را در مرحله بعد انتخاب می‌کند. مهم‌ترین معایب این سیستم‌ها این است که برای شبیه‌سازی یک سبک موسیقی، باید قطعه‌های بسیاری از همان سبک تحلیل و احتمالات پیدا شوند و دیگر اینکه درنظرگرفتن معیارهای سـطح بالاتر موسیقی می‌تواند موجب پیچیدگی وصف ناپذیر الگوریتم و همینطور دشواری پیاده‌سازی الگوریتم گردد.

۲) رویکردهای GAN[1]:

روش دیگر، استفاده از رویکردهای GAN می باشد که در آن توانایی تولید خودکار مولدهای ملودی توسط یک منتقد ماشین هدایت می‌شود. SeqGAN و MuseGAN برنامه‌های تولید شبکه‌های توالی هستند [11, 12] و موسیقی متن‌های چند صدایی را تشکیل می دهند [13]. در این روش‌ها، مولدهای ملودی هنوز شبکه‌های عصبی تکراری هستند که پیروی از آن‌ها دشوار است. مشکل دیگر این است که مولدهای بازگشتی در بیان تغییرات بلندمدت بی‌اثر هستند.

۳) یادگیری ماشین (ML[2]):

MLها در این زمینه زیاد استفاده شده‌اند. ویدمور در سال ۱۹۹۲ برای هارمونیزی کردن ملودی‌ها ازML استفاده کرد. با اینکه جالب است سیستمی از طریق مثال یاد بگیرد اما این یادگیری در موارد بسیاری نیاز به پالایش دارد زیرا مثال‌های متعددی می‌توانند باهم تضاد داشته باشند [10, 14] .

۴) سیستم‌های هیربد:

این سیستم‌ها از روش‌های مختلف هوش مصنوعی استفاده می‌کنند. سیستم‌های هیربد معمولا از پیچیدگی‌های بالایی برخوردارند. مخصوصا در مواردی که سیستم‌ها به سختی باهم ترکیب شده باشند. در نهایت، عیب این سیستم‌ها این است که با آهنگسازی به عنوان یک کار حل مساله برخورد می‌کنند تا مساله‌ای که با خلاقیت همراه باشد.

در نهایت در مقاله ای که Zhiyang Xiang و Yibo Guo در سال ۲۰۲۰ منتشر کردند[16] یک روش تولید موسیقی خودکار مبتنی بر هوش مصنوعی که در آن مولد ملودی یک ترکیب گاوسی است، پیشنهاد کردند؛ که در آن مداخله ساده‌تر است و حد طول CNN[3] را ندارد. همچنین این ساختارها از سازه‌های کاملا مشخص ساخته شده‌اند و موسیقی متن را براساس انتقال الگوهای ساخته شده، شکل می‌دهند.

۴– مدل مخلوط گاوسی (GMM[4])

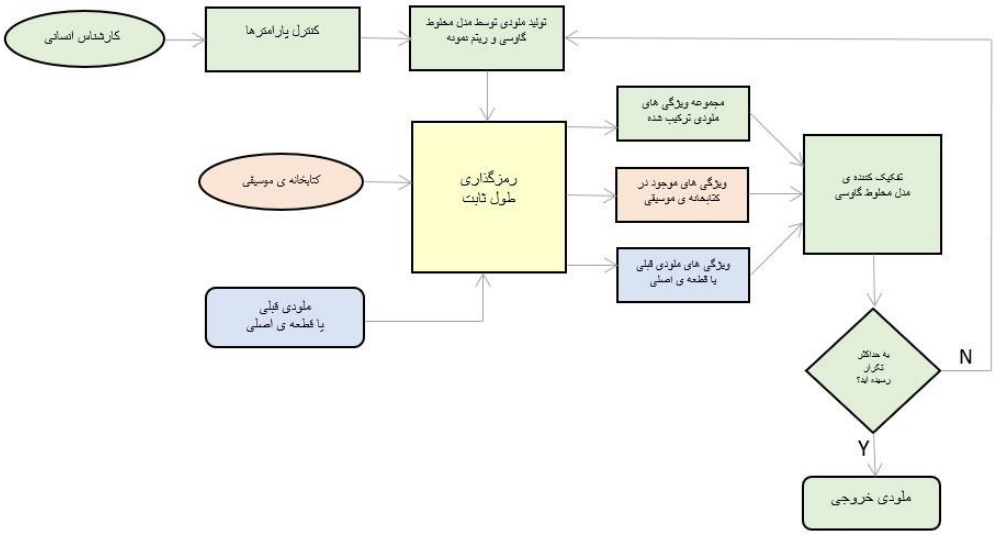
یک تابع چگالی احتمالاتی پارامتری است که مجموع وزن‌دار چگالی مولفه‌های گاوسی را نمایش می‌دهد. این مدل جهت بیشینه کردن شاخصه درست‌نمایی یا احتمال، از تثبیت برخی پارامترها در هر بار تکرار استفاده می‌کند. پارامترهای مربوط به

[1] Generative Adversarial Network
[2] Machine Learning
[3] Convolutional Neural Network
[4] Gaussian Mixture Model

آن با استفاده از داده‌های پیشینه سازی آموزش و به‌کارگیری الگوریتم تکرار EM[1] (امید ریاضی) تخمین زده‌میشوند که تخمینی از مدل ابتدایی به وسیله یک داده آموزشی ایده آل است.

۴-۱- استفاده از مدل مخلوط گاوسی در تولید خودکار موسیقی

در این روش، موسیقی متن را براساس تولید الگوها و انتقال الگوها می‌سازند. ابتدا چارچوبی برای نمایش موسیقی جدید طراحی شده است تا ملودی‌های با طول متغیر به بردارهای با طول ثابت تبدیل شوند. سپس یک مدل مخلوط محدود گاوسی برای مولدهای ملودی قابل کنترل ساخته شده است که مداخلات انسان به عنوان محدودیت در فضای پارامتر پنهان مدل سازی می‌شود. در نهایت، یک فرآیند GAN برای نظارت خودکار بر روند تولید گنجانیده شده است که لایه دیگری از GMM به عنوان یک تفکیک کننده برای هدایت مولد GMM اضافه می‌شود. مولد ملودی GMM قادر است همزمان از دستورالعمل‌های خودکار انسان و AI پیروی کند. چارچوب کامل در شکل۱ نشان داده شده است. نت‌های موسیقی توسط mingus به بردارهای عددی ترجمه می‌شوند [15].



شکل۱. الگوریتم استفاده از مدل مخلوط گاوسی در تولید موسیقی

یادگیری GAN شامل آموزش یک تولیدکننده و یک تفکیک‌کننده است. که مورد دوم با ویژگی‌های ملودی "واقعی" تولیدشده توسط کتابخانه موسیقی موجود و ویژگی‌های ملودی "ساختگی" از مولد آموزش داده می‌شود. در این کار دو فرآیند مستقل GAN طراحی می‌شود ؛ یکی از آن‌ها مسئول تولید ملودی است و دیگری مسئول انتقال ملودی است. GAN انتقال نه

[1] Expected value

تنها با انتقال‌های "ساختگی" در تکرار فعلی، بلکه باقیمانده‌های انتقال‌های قبلی نیز آموزش داده می‌شود تا به یک تداوم در مسیر کامل تولیدشده دست یابد. از آنجا که متغیرهای مولد به طور مداوم تنظیم می‌شوند ، تفکیک‌کننده از توانایی ترکیب ویژگی‌های ایجادشده جدید و فراموشی ویژگی‌های قدیمی برخوردار است. در [5] یک چارچوب آموزشی افزایشی برای تفکیک‌کننده‌ی GMM طراحی می‌شود. زیرآهنگ‌ها به عنوان انتقال الگو از مسیر اصلی درنظر گرفته می‌شوند ، بنابراین ترکیبات چند صدایی از همان چارچوب ترکیب تک آهنگ پیروی می‌کنند. به عبارت دیگر انتقالات الگو، به خوبی تولیدکنندگان زیرآهنگ، سازمان دهنده‌ی ملودی هستند.

یک روش یادگیری با نمایش ثابت با جزئیات دقیق در این بخش ارائه شده است. این چارچوب امکان استفاده از بسیاری از روش‌های تجزیه و تحلیل الگوی سنتی (از جمله GMM) را در کار تولید موسیقی فراهم می‌کند. چندین نت به‌طور همزمان به عنوان ترک‌های زیرآوایی فرض می‌شوند که قرار است به‌طور مستقل از آهنگ(ترک) اصلی کدگذاری شوند. بعد از تفکیک ترک‌ها، یک ملودی در یک آهنگ آوایی، از نت‌ها و مدت‌ها (ضرب‌ها) تشکیل شده است که می‌تواند به صورت دو نواخت(tuple) نمایش داده‌شود. یک نواخت، شامل نت‌ها و شاخص‌های شروع آن‌ها و نواخت دیگر، شامل نت‌ها و شاخص‌های توقف آن‌ها است. از آنجا که وظایف مدل سازی دو نواخت مشابه یکدیگر هستند، کار رمزگذاری نیز به شرح زیر ساده می‌شود. شاخص زمان شروع یک نت، شاخص زمان توقف نت قبلی است و نت‌هایی که همزمان پخش می‌شوند به زیرآهنگ‌ها تقسیم می‌شوند. در نتیجه، ملودی در آهنگ‌های اصلی به نت‌ها و مدت زمان طول کشیدن آن‌ها ساده می‌شود.

فرض می‌کنیم یک ملودی، مشاهده شونده‌ای $(observation)$ از تابع $m(t) \in R$ است که در بازه زمانی $[-1,1]$ تعریف شده است، و دنباله $\{ti\}$ با عناصری در بازه‌ی $[-1,1]$ که شاخص‌های زمان شروع نت‌ها است، یک ناظر $(observer)$ می‌باشد. در حقیقت مشاهده‌شده $(observed)$ $\{ni|ni = m(ti)\}$ مجموعه نت‌ها می‌باشد [5].

۴-۲- استفاده از چارچوب اصلاح شده ی GAN

در این بخش، یک چارچوب GAN اصلاح شده برای تولید موسیقی متن از لیستی از پارامترهای کنترل fcg و تعداد نت‌ها پیشنهاد می‌شود. دو تفکیک‌کننده‌ی ماشین به جای یک مورد در GAN اصلی، با رمزگذاری طول ثابت ملودی‌ها و انتقال‌ها آموزش می‌بینند. این اصلاح برای آموزش‌های آسان افزایشی و کاهشی تفکیک‌کننده‌هاست. سپس، ناظران توابع ملودی که شاخص‌های زمانی تولیدشده توسط مجموعه‌های مدت زمان هستند، از تعداد نت‌های مشخص شده انسان در هربار تولید، محاسبه می شوند. هنگامی که تعداد نت از حداکثر تعداد نت در یک ملودی کوچکتر است و حداکثر شمارش نت توسط کل نت‌های ربع یک قطعه ملودی قابل نگهداری است و حداقل مدت زمان تعریف شده است، یک شمارش نت می تواند تعداد زیادی از مجموعه‌های مدت(ضرب) ممکن را ایجاد کند. سرانجام، قطعه‌های موسیقی با توابع ملودی $m(t)$ به عنوان مولد اصلی و مجموعه‌های مدت زمان به عنوان نمونه ساخته می‌شوند. همچنین دو تابع تفکیک‌کننده‌ی GMM در نقش یک ماشین تمیزدهنده برای انتخاب بهترین ترکیبات عمل می‌کنند.

روش فرآیند گاوسی (GP[1]) برای انجام این کارهای طبقه‌بندی انتخاب شده است، زیرا عملکرد آن در مشکلات بعدی(dimensional) کم است. هر قطعه از ملودی اغلب بیش از ۳۰ نت ندارد و در مقایسه با موسیقی متن کامل و سایر رسانه‌ها مانند اسناد متنی، ابعاد کمی دارد. در این روش با الهام از کارهای قبلی، تقریب توزیع فرآیندهای گاوسی طراحی می‌شود که می‌تواند به صورت فزاینده به‌روز شود [16]. یک فرآیند گاوسی که یک کلاس از نمونه‌ها را مدل‌سازی می‌کند می‌تواند با چندین روش آموزش مدل مخلوط گاوسی به صورت تدریجی به‌روز شود ، اما تقریب تدریجی فرآیندهای گاوسی که دو کلاس را مدل می‌کنند، دشوار است. این بدان دلیل است که فرآیندهای گاوسی طبقه‌بندی از یک تابع نهفته برای تخمین احتمالات مرتبط با ورودی مربوط به کلاس استفاده می‌کنند[17] و تغییرات مجموعه آموزش از به‌روزرسانی تدریجی آسان این تابع نهفته جلوگیری می‌کند. به دلیل دشواری آموزش تدریجی فرآیندهای گاوسی برای طبقه‌بندی، در جایی که دو تفکیک وجود داشته باشد، معادل چارچوب GAN طراحی می‌شود.

۴-۳- بررسی آزمایشات

در این بخش، توانایی‌های کنترل چارچوب ذکرشده نشان داده می‌شود و با استفاده از منتقدان انسانی، نتایج با روش‌های موجود مقایسه می‌شود[5]. طرفداران پرونده‌های MIDI[2] سه‌گانه Ninja Gaiden NES را به عنوان کتابخانه موسیقی انتخاب کرده‌اند. این انتخاب دو دلیل دارد:
اول اینکه این موسیقی متن‌ها از نظر ژانرهای موسیقی شبیه یکدیگر هستند. دوم، این مخزن بزرگترین کتابخانه موسیقی بازی است که به صورت رایگان در قالب MIDI در دسترس است. در دو ترکیب نهایی، موسیقی متن در ژانرهای مختلف نیز در آموزش‌ها گنجانده شده است، تا نشان دهد که توانایی ترکیبات ساختار طولانی‌مدت تحت تأثیر کتابخانه‌های آموزشی نیست. نمونه‌های آموزش اضافی از انجمن Musescore بارگیری می‌شود. GMM ها با استفاده از زبان برنامه‌نویسی پایتون و پکیج pygks اجرا می‌شوند [18].

برای اجرای کنترل بلندمدت و کوتاه‌مدت از سیستم دوسلسله‌مراتبی استفاده می‌شود. پارامترهای کنترل شامل سه آرایه هستند. در سلسله مراتب پایین، آرایه‌ای از اعداد صحیح انتقال‌های غیر ملودی را کنترل می‌کند، که سطح متوسط ارتفاع قسمت بعدی یک قطعه ملودی واحد را که باید بالاتر از قسمت قبلی باشد، تعریف می‌کند. در سلسله‌مراتب بالاتر، آرایه دیگری انتقال بین ملودی‌های واحد را کنترل می‌کند و سپس اینکه میانگین سطح صدا در هر قطعه ملودی باید بالاتر از قطعه قبلی باشد، را تعریف می‌کند. سومین آرایه کنترل‌کننده تعداد نت‌ها را در هر ملودی واحد تعریف می‌کند. از آنجا که ما بیش از مهارت‌های بازی یا تعدیل، روی سطوح و ریتم‌های سطح تمرکز می‌کنیم، نت‌های متوالی با همان سطح صدا به عنوان یک نت واحد درنظر گرفته می‌شوند تا روند انتخاب ریتم ساده شود. در همه آزمایش‌ها، مدت زمان یک ملودی واحد چهار ربع نت است و یا به‌عبارتی در سیستم نت‌نویسی ۴/۴ به شکل دولاچنگ پیاده‌سازی شده است.

آزمایش‌ها شامل چهار قسمت است. درابتدا، عملکرد تک‌ملودی ارزیابی می‌شود. دوم، توانایی کنترل سطح گام با اعمال تغییرات سطح گام در مسیرهای طولانی‌تر نشان داده می‌شود. سوم، کنترل ریتم در بالای کنترل سطح گام اضافه می‌شود. چهارم، یادداشت های همزمان از طریق آهنگ های زیر آهنگ اضافه می شوند. زیر آهنگ ها با روش انتقال الگو اضافه

[1] Gaussian Processes
[2] Musical Instrument Digital Interface

می‌شوند که همان چارچوب برای آهنگسازان اصلی ملودی است. آزمایشات مقایسه‌شده توسط منتقدان انسانی ارائه شده است؛ جایی که موسیقی متن ساخته شده بدون کنترل انسان با GAN رقابت می‌کند. در این آزمایش یازده قطعه موسیقی بدون هیچ‌گونه انتخاب یا ویرایش در این مسابقه تولید شده است. کیفیت موسیقی متن تولید شده به‌طور تصادفی مستقیما توسط منتقدان انسانی رأی‌گیری شده است. گروه منتقدین درمجموع ۱۶ بازیکن را شامل می‌شوند، ۴ نفر از آن‌ها آموزش موسیقی و ۱۲ نفر هیچ‌گونه آگاهی از ترکیب موسیقی متن ندارند.

برای نشان دادن ساختار طولانی‌مدت موسیقی متن، آنتروپی قطعات اصلی موسیقی کلاسیک (چهارمین سمفونی چایکوفسکی)، موسیقی پاپ (قهرمان کوتاه تغییر توسط هوی)، موسیقی کانتری‌رود (Country Road توسط جان دنور)، جاز (Galaxia توسط Gylfie)، موسیقی الکترونیکی (Clubbed to Death توسط Rob Dougan)، آهنگ‌های صدای بازی، آهنگ‌های تولیدشده توسط RNN، LSTM، آهنگ‌های CNN (MuseGAN و MidiNet) و آهنگ‌های تولیدشده GMM-GAN مورد استفاده قرار گرفته‌اند.

از بررسی‌های انجام‌شده می‌توان دریافت که آهنگ‌های صوتی بازی کاملا ساختاریافته هنگام حرکت به جلو، دارای چندین افت آنتروپی هستند. این کاهش آنتروپی حداقل به دو دلیل ایجاد می‌شود. اول، آهنگ‌های بازی اغلب تکرارمی شوند که وقتی زمان به جلو می‌رود، بازیکنان برای درک موسیقی متن به توجه کمتری نیاز دارند. دوم، آهنگ‌های بازی حاوی اطلاعات بی پایان نیستند و بین چندین ملودی اصلی عبور می‌کنند.

موسیقی کلاسیک، پاپ و کانتری با ساختارهای طولانی مدت با رسیدن به حداکثر امتیاز روند آنتروپی کاهش می‌یابد. فواصل خاموش باعث می‌شود منحنی‌های آنتروپی موسیقی کلاسیک از سایرین نرم‌تر باشد. GMM-GAN تولید "هشت و نیم" و "نهم" با ساختارهای پیچیده دارای منحنی آنتروپی مشابه موسیقی کلاسیک پس از حداکثر امتیاز است. «tenth.mid» و «eleventh.mid» از کتابخانه آموزش میکس تولید می‌شوند و منحنی‌های آنتروپی آن‌ها شبیه به موسیقی کلاسیک، کانتری و موسیقی پاپ است. این بدان معنی است که ساختار کنترل‌شده طولانی‌مدت تحت تأثیر کتابخانه‌های آموزشی نیست. مسیرهای تولید شده توسط RNN، LSTM و CNN تمایل به افزایش آنتروپی در هنگام طولانی بودن مسیرهای تولید شده دارند، که باعث می شود رفتارهای طولانی مدت آنها شبیه مولدهای تعداد تصادفی باشد. به طور خلاصه، نتایج توانایی چارچوب مطرح شده در تحمیل ساختارهای طولانی مدت به موسیقی متن تولید شده را نشان می‌دهد. در آزمایش نهایی، روش پیشنهادی بدون کنترل و روش‌های RNN و CNN به GAN مقایسه می‌شود.

۵– ارزیابی کلی

برخی از منتقدان اظهارنظر کرده‌اند که RNN و CNN ملودی‌های روان تولید می‌کنند، اما به‌نظر می‌رسد برای بازی در یک رایانه خیلی ساده است. ملودی‌های تولیدشده توسط RNN تقریباً در همان سبک موسیقی هستند. از نتایج نظرسنجی، مشخص است که چارچوب پیشنهادی قادر به ساخت موسیقی خودکار است که از ساختارهای ساده تعریف شده توسط انسان پیروی می‌کند و در ترکیب موسیقی متن بازی بهتر از مولد RNN است.

به طور خلاصه، چارچوب پیشنهادی دارای مزایای متعددی نسبت به روش‌های اصلی است و در عین حال چند مشکل برای حل و فصل دارد. دو اشکال عمده وجود دارد: اول، سیستم آهنگسازی به دلیل روند نمایش عددی فقط بر نت‌ها و ریتم‌ها تمرکز می‌کند. دوم، بهینه‌سازی‌های عددی فقط با روش‌های شیب‌دار انجام می‌شود. همچنین سه مزیت اصلی وجود دارد: اول،

یک مکانیزم کنترل سلسله‌مراتبی باعث تعادل ساختارهای بلندمدت و کوتاه‌مدت می‌شود. دوم، مکانیسم‌های کنترل ساده هستند و آهنگسازی توسط افراد و بدون هیچ‌گونه تخصص موسیقی انجام می‌شود. سوم، در چارچوب ترکیب خودکار، یک روش نمایش موسیقی GMM وجود دارد که می‌تواند درکنار بسیاری از تکنیک‌های بهینه‌سازی محدود دیگر برای حل مسئله ترکیب موسیقی AI با هدف کلی استفاده شود.

۶– نتیجه گیری

به‌طور کلی RNNها به‌طور گسترده‌ای برای مشکلات تولید موسیقی رایانه استفاده می‌شوند. آن‌ها در بداهه‌پردازی مهارت دارند و به‌صورت متناوب ملودی‌های جالبی تولید می‌کنند اما در پیروی از ساختارهای بلندمدت معنی‌دار خروجی مناسبی از خود نشان می‌دهند. ویژگی‌های کنترل‌نشده RNN، کارایی ساخت هوش مصنوعی موسیقی متن بازی را باطل می‌کند. یک چارچوب ترکیب موسیقی اتوماتیک در این مقاله مورد مطالعه قرار گرفته است، جایی که ملودی ها به جای سیستم های عملکرد تکرار غیرخطی در آثار سنتی به عنوان GMM مدل سازی می‌شوند. ما محدودیت‌های سطح گام را برای GMM ها و انتقال GMM ها اعمال می‌کنیم تا ساختارهای کوتاه‌مدت و بلندمدت را برای رفع اشکال مولدهای موسیقی RNN اعمال کنیم. مشکلات انتخاب تعداد یادداشت‌ها به عنوان فرآیندهای نمونه‌گیری در چارچوب GAN حل می‌شوند. روش پیشنهادی قادر به ترکیب موسیقی چند صدایی است که از ساختارهای تعریف شده توسط انسان پیروی می‌کند، در حالی که دانش کاربر از موسیقی نیازی به تخصص ندارد.

رندر خودکار پرونده‌های MIDI یکی دیگر از عناصر پوسته‌ای است که هم عملکرد و هم کارایی ترکیب هوش مصنوعی را افزایش می‌دهد. در کارهای بعدی، تکنیک های ارائه MIDI با چارچوب GAN قابل بررسی هستند. علاوه براین، یادگیری ساختار جهانی خودکار براساس چارچوب پیشنهادی، درجه اتوماسیون را در ترکیب موسیقی متن بازی بیشتر می‌کند. در حالی که بازیکن در حال تجربه توطئه‌های بازی یا غوطه‌ورشدن در محیط‌های مجازی است، روش‌های هوش مصنوعی توانایی آهنگسازی در زمان واقعی را دارند

۷–مراجع

1. *Poole, D., A. Mackworth, and R. Goebel, Computational Intelligence. 1998.*
2. *Karpov, N., Artificial Intelligence for Music Composing: Future Scenario Analysis. 2020.*
3. *Tsiros, A. and A. Palladini, Towards a Human-Centric Design Framework for AI Assisted Music Production.*
4. *Koh, A., Music for AI Reports: Dual Prospects in Music Production. 2018.*
5. *Xiang, Z. and Y. Guo, Controlling melody structures in automatic game soundtrack compositions with adversarial learning guided Gaussian mixture models. IEEE Transactions on Games, 2020.*
6. *Manovich, L., AI aesthetics. 2018: Strelka Press Moscow.*
7. *Mozer, M.C., Neural network music composition by prediction: Exploring the benefits of psychoacoustic constraints and multi-scale processing. Connection Science, 1994.* ***6****(2-3): p. 247-280.*
8. *Hochreiter, S. and J. Schmidhuber, Long short-term memory. Neural computation, 1997.* ***9****(8): p. 1735-1780.*
9. *Huang, C.-Z.A., et al. Music transformer: Generating music with long-term structure. in International Conference on Learning Representations. 2018.*

10. *Papadopoulos, G. and G. Wiggins. AI methods for algorithmic composition: A survey, a critical view and future prospects. in AISB symposium on musical creativity. 1999. Edinburgh, UK.*
11. *Dong, H.-W., et al. Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment. in Proceedings of the AAAI Conference on Artificial Intelligence. 2018.*
12. *Lee, S.-g., et al., Polyphonic music generation with sequence generative adversarial networks. arXiv preprint arXiv:1710.11418, 2017.*
13. *Yu, L., et al. Seqgan: Sequence generative adversarial nets with policy gradient. in Proceedings of the AAAI conference on artificial intelligence. 2017.*
14. *Nierhaus, G., Algorithmic composition: paradigms of automated music generation. 2009: Springer Science & Business Media.*
15. *Ramsay, A., Modelling Classical Music with Machine Learning. 2016.*
16. *Xiang, Z., et al., Semi-parametric training of autoencoders with Gaussian kernel smoothed topology learning neural networks. Neural Computing and Applications, 2020.* ***32****(9): p. 4933-4950.*
17. *Kuss, M. and C.E. Rasmussen, Assessing approximate inference for binary Gaussian process classification. Journal of machine learning research, 2005.* ***6****(Oct): p. 1679-1704.*
18. *Xiang, Z., et al., Gaussian kernel smooth regression with topology learning neural networks and Python implementation. Neurocomputing, 2017.* ***260****: p. 1-4.*