6th Conference on Recent Research in Science and Technology
ششمین همایش پژوهش های نوین در علوم و فناوری
تابستان ۱۳۹۵
Summer 2016

پردازش زبان طبیعی و بازیابی اطلاعات موسیقایی

ناهید امیری

دانشجوی دکتری علم اطلاعات و دانش شناسی دانشگاه خوارزمی

nahidamirilib@gmail.com

Archive of SID

چکیده

هدف این مقاله بررسی کارکردهای پردازش زبان طبیعی در بازیابی اطلاعات موسیقایی است. پژوهش‌ها در حوزه موسیقی محاسباتی نشان می‌دهد که حوزۀ پردازش زبان طبیعی به پژوهش‌های جدید صورت گرفته در موسیقی نزدیک است. بنابراین می‌توان از این شباهت سود جست و تکنیک‌های پردازش زبان طبیعی را در موسیقی به کار بست. در این مقاله شباهت میان هر دو حوزۀ فوق توصیف می‌شود. همچنین سطوح پردازش زبان طبیعی و کارکرد موسیقی محاسباتی تبیین می گردد و در ادامه به نقاط مشترک پژوهش های بازیابی اطلاعات موسیقایی و حوزۀ پردازش زبان طبیعی اشاره می‌شود. در پایان، پژوهش‌های انجام شده دربارۀ مدل‌سازی موسیقی آماری با نظریات مشابه در حوزۀ پردازش زبان طبیعی مقایسه و رهنمون‌هایی برای بازیابی اطلاعات موسیقایی ارائه می‌گردد.

کلمات کلیدی: بازیابی اطلاعات، پردازش زبان طبیعی، بازیابی اطلاعات موسیقایی

۱. مقدمه

همراه با انقلاب اطلاعاتی که با ورود رایانه‌ها شکل گرفت، فرصت‌های نوینی برای هنرمندان و پژوهشگران موسیقی نیز ایجاد شد. با اجرای مسائل محاسباتی بر روی پردازنده‌های مرکزی بزرگ رایانه‌ای، از آنها برای ایجاد موسیقی محاسباتی نخستین استفاده شد. در همان زمان، بررسی دربارۀ چگونگی استفاده از رایانه‌ها برای پردازش و تحلیل مواد موسیقایی آغاز شد. موسیقی، شبیه به گفتار انسانی، از ابتدای تکامل انسان با او همراه بوده است، به‌نحوی‌که فهم عمیق موسیقی می‌تواند فهم بهتر شناخت انسانی را ممکن سازد. با وجود این، هنوز داده‌های موسیقایی در بیشتر موارد به عنوان داده‌های دوتایی درهم‌ریخته‌ای قلمداد می‌شوند که در ردیف تصاویر، فیلم‌ها و برنامه‌های رایانه‌ای قرار دارند؛ برخلاف داده‌های متنی که پردازش، جستجو و نمایه‌سازی آنها آسان است، با تعداد زیادی رایانۀ دردسترس ایجاد می‌شوند و تکنیک‌های عرضه‌شده از سوی NLP، به بازیابی اطلاعات یا استخراج داده‌های متنی همچون طبقه‌بندی، تحلیل، خلاصه‌سازی، نمایه‌سازی، جستجو، ترجمه و بسیاری امور دیگر کمک می‌کنند. با وجود این، می‌توان موسیقی را نیز به عنوان یک زبان طبیعی قلمداد کرد و آن را به‌روشی مشابه متن پردازش نمود. اگرچه تفاوت‌های اساسی میان متن مکتوب و موسیقی وجود دارد، اما آنها ویژگی‌های مشترک زیادی نیز دارند(Wołkowicz, 2010, p.665). بخش قابل‌توجهی از اصوات موسیقایی، ترانه‌ها هستند. امروزه علاقۀ زیادی برای دستیابی

www.SID.ir

به محتوای موسیقایی وجود دارد. کاربران علاوه بر جستجو برای داده‌های ویراستاری، همچون هنرمند یا نام آهنگ، می‌توانند در طبقه‌بندی‌های ژانری که افراد مختلف انجام‌داده‌اند، نیز بگردند یا از پیشنهادات سیستم‌های فیلترینگ گروهی پیروی‌کنند. فیلترینگ گروهی از این گونه اطلاعات استفاده می‌کنند: «افرادی که به X گوش داده‌اند، به Y نیز تمایل داشته‌ان (Mahedero,2005, p.1). در این مقاله برخی از وظایف NLP در حیطهٔ ترانه‌ها مطرح می‌گردد. این وظایف عبارتند از شناسایی زبان، کشف ساختار و مقوله‌بندی متن.موسیقی موضوع فراگیری در جوامع انسانی است، چراکه تقریباً همگان به‌نحوی به موسیقی گوش می‌دهند و افراد بسیاری نیز آن را می‌آفرینند. داوین[1]معتقد است که پژوهش در حوزهٔ بازیابی اطلاعات موسیقایی نخست با بحث در خصوص استخراج و استنباط مشخصه‌های معنی‌دار از موسیقی (از سیگنال‌های صوتی، بازنمودهای نمادین یا منابع خارجی همچون صفحات وب)، نمایه‌سازی موسیقی با استفاده از این مشخصه‌ها و گسترش تمهیدات مختلف برای *جستجو و بازیابی اطلاعات* (برای نمونه، جستجوی محتوابنیاد، سیستم‌های پیشنهادکنندهٔ موسیقی، یا واسط‌های کاربری برای کاوش در مجموعه‌های بزرگ موسیقی) آغاز شد. در نتیجه، بازیابی اطلاعات موسیقایی[2] در پی این هدف است که بازار گسترده و وسیع موسیقی جهان را در دسترس عموم مردم قرار دهد. برای رسیدن به این هدف، بازنمودهای مختلف اشخاص مرتبط با موسیقی (مثلاً ترانه‌سرایان، آهنگساران، نوازندگان و فروشندگان) و موضوعات مرتبط با موسیقی (مثلاً قطعات موسیقایی، آلبوم‌ها و ویدئو کلیپ‌ها) مورد بررسی قرار می‌گیرند.بازیابی اطلاعات موسیقایی، یک حوزه پژوهشی میان‌رشته‌ای است که به استخراج، تحلیل و استفاده از اطلاعات دربارهٔ موسیقی و صوت می‌پردازد. به‌طورمرسوم، بازیابی اطلاعات موسیقایی بیشتر بر استفاده از محتوای صوتی برای براوردکرن اطلاعات بافتی متمرکز بوده‌است (Schedl, 2014, pp. 205-213).

۲.موسیقی به عنوان زبان طبیعی

زبان طبیعی، بنا به تعریف، عبارت است از هر زبانی که به عنوان نتیجهٔ استعداد فطری عقل انسان برای زبان به شکلی از پیش فکرنشده ایجاد می‌شود. همگان در اینکه موسیقی در این تعریف می‌گنجد، توافق ندارند، اما پژوهشگران موسیقی، که با قواعد موسیقایی آشنا هستند، تمایل بیشتری برای پذیرش آن دارند. موسیقی، همانند متن از بازنمودی نمادین برخوردار است که منشاء آن به دوران باستان بازمی‌گردد. موسیقی و زبان تنها فعالیت‌های خلاقانهٔ انسان هستند که بازنمودهای نمادین به‌طور رایجی در آنها استفاده می‌شود. دیگر فعالیت‌ها، همچون نقاشی، مجسمه‌سازی و رقص چنین نت‌نویسی نمادین مشترکی ندارند. نت‌نویسی موسیقی را نمی‌توان همانند متن به رایانه منتقل ساخت، اما به راحتی می‌توان به این مسأله فائق آمد. برای نمونه، در برابر این استدلال که متن را می‌توان به آسانی به کلمات تقسیم کرد، که یکی از ویژگی‌های اولیهٔ NLP و بازیابی اطلاعات است، اما این در مورد موسیقی صدق نمی‌کند، می‌توان با توجه به این امر پاسخ داد که زبان‌های طبیعی‌ای، همچون زبان تایلندی، وجود دارند که از چیزی برای تقسیم کلمات استفاده نمی‌کنند.بازیابی اطلاعات موسیقایی موضوعی میان‌رشته‌ای است، زیرا از سویی شامل عناصر اساسی موسیقی و از سوی دیگر دربردارندهٔ عناصری از علم اطلاعات است. اولین اثری که در این حوزه نوشته شد، در نیمهٔ دههٔ ۱۹۶۰ به دست مایکل کسلر[3] و دیگران انجام گرفت(Kassler, 1970). مقایسهٔ موسیقی و متن، امری راهگشا و مفید است (Byrd, 2002, p.5).

بیشینه	میانه	کمینه	ساختار صریح

[1] . Dowine
[2] . Music Information Retreival (MIR)
[3] . Michael Kassler

Archive of SID
www.SID.ir

نت‌نویسی موسیقایی (برگه نت)	رخدادها (فایل MIDI استاندارد)	صوت (CD، MP3)	بازنمود موسیقایی
متن HTML	متن معمولی	صوت (گفتار)	بازنمود متنی

جدول۱. مقایسه موسیقی و متن

تقریباً همهٔ انواع موسیقی که به گوش ما آشنا هستند، از نت‌ها ساخته‌می‌شوند همان‌گونه که متن‌ها از حروف و کلمات ساخته می‌شوند. نت‌ها بیشتر شبیه حروفند تا کلمات، هرچند این شباهت چندان روشن نیست.

چهار پارامتر اصلی نت‌های موسیقایی عبارتند از: نواک یا ارتفاع صوت[1]: زیر و بمی صدا است. درکی است که گوش شنونده از بسامد یا زیر و بم بودن نت‌ها دارد.دیرند[2] : کشش ارزش زمانیِ نت‌های موسیقی.بلندی[3]: اثر شدت صوت در شنوایی. طنین[4]: کیفیت صدایی که از ساز موسیقی به‌خصوصی به وجود می‌آید.انسانها موسیقی را به‌شکلی غیر خطی می‌شنوند(Pierce, 1992, pp. 211-212).برخی موسیقی‌ها تک‌صدایی هستند یعنی در هر لحظه تنها صدای یک نت وجود دارد. اما بیشتر موسیقی‌های غربی چندصدایی هستند، یعنی در هر لحظه چندین نت به گوش می‌رسد. چندصدایی بودن موسیقی، بازیابی اطلاعات موسیقایی را دشوارتر می‌سازد. یکی از مشکلات بازیابی در حوزه موسیقی در خصوص موسیقی‌های چندصدایی مسألهٔ «برجستگی» است، یعنی مسأله تشخیص میزان اهمیت یک عنصر در موسیقی، چه این عنصر نت باشد و چه آکورد، ملودی یا هرچیز دیگر. تکنیک‌های NLP فراوانی در بازیابی اطلاعات به کاربرده‌شده‌اند، اما نتایج چندان موفقیت‌آمیز نبوده است (Brants, 2015, p.1).

۳. وظایف پردازش زبان طبیعی

برخی از مهم‌ترین وظایف NLP به شرح زیر است(Nadkarni, 2011, pp.544-549):

۳-۱.تحلیل ارجاع: با فرض یک جمله یا بخش بزرگتری از یک متن، تعیین اینکه چه کلماتی به امور واحدی ارجاع دارند، تحلیل ارجاع خوانده می‌شود. نمونهٔ خاص این وظیفه، بررسی anaphora یعنی مرجع‌داری، است.در زبان‌شناسی به حالتی در جمله‌ها که درک معنای یک عنصر متنی با مراجعه به عناصر دیگر متن امکان‌پذیر می‌شود ارجاع می‌گویند. عمده ارجاعات را ضمایر شخصی، ملکی، اشاره‌ای و صفات ملکی تشکیل می‌دهند.

۳-۲. تحلیل گفتمان: گفتمان‌شناسی یا تحلیل گفتمان یا تحلیل کلام اصطلاحی کلی برای اطلاق به مطالعاتی است که زبان نوشتاری، گفتاری یا نشانه‌ای یا هر گونه پدیده نشانه‌شناختی را مورد تجزیه و تحلیل قرار می‌دهند. تحلیل گفتمان معمولاً یکی از زیرشاخه‌های علم زبان‌شناسی شناخته می‌شود.

۳-۳.ترجمهٔ ماشینی: ترجمهٔ ماشینی زیر شاخه‌ای از زبان‌شناسی محاسباتی است که عبارت است از ترجمهٔ متنی از یک زبان طبیعی به زبانی دیگر، توسط کامپیوتر. در سطح مقدماتی، ترجمه ماشینی یک جایگزینی ساده برای کلمات از زبان طبیعی به زبان دیگری است. با استفاده از تکنیک‌های زبان‌شناسی پیکره‌ای، ترجمه‌های پیچیده بیشتری قابل دستیابی هستند. همچنین این تکنیک‌ها کنترل بهتر تفاوت‌های گونه‌شناسی در زبان، تشخیص عبارات و ترجمهٔ اصطلاحات را به خوبی و درستی جدا کردن عبارات نامتعارف در متن، مقدور می‌سازند.

[1] . Pitch
[2] . Duration
[3] . loudness
[4] . Timbre

Archive of SID
www.SID.ir

۳-۴. تقطیع صرفی: تقطیع کلمات به تکواژها و مشخص کردن طبقهٔ آنها. کوچک‌ترین یکای (واحد) زبان که دارای نقش دستوری و معنایی مستقل می‌باشد تکواژ یا علامت نام دارد. هجاها و طول آنها نمی‌توانند برای بازشناختن تکواژها ابزار سودمندی باشند. معیار و سنجهٔ بنیادین این است که تکواژ را نمی‌توان به یکاهای دستوری کوچک‌تر بخش کرد. دشواری این وظیفه بیشتر بر پیچیدگی صرف‌شناختی (ساختار کلمات) زبان مورد بحث بستگی دارد.

۳-۵.تشخیص هستارهای اسمی: تعیین اینکه در جریان یک متن، هریک از آیتم‌ها با کدام یک از اسامی خاص، مثل افراد یا مکانها، ارتباط دارد و هر یک از این اسامی ذیل کدام گونه قرار دارد (مثلاً شخص، مکان یا سازمان). لازم به ذکر است که اگرچه در برخی از زبان‌ها مثل زبان انگلیسی، حرف بزرگ در ابتدای کلمه می‌تواند به تشخیص هستار اسمی کمک کند، اما این اطلاعات نمی‌تواند در تشخیص گونهٔ هستار اسمی کمکی کند، و غالباً نیز نادقیق یا ناکافی است. چرا که مثلاً کلمهٔ اول جمله نیز با حرف بزرگ آغاز می‌شود. همچنین، هستارهای اسمی کلمات گوناگونی را در بر می‌گیرند که فقط برخی از آنها با حروف بزرگ آغاز می‌شوند. افزون بر این، بسیاری از زبان‌های دیگر مثل چینی یا عربی اصلاً از حروف بزرگ در ابتدای اسامی خاص استفاده نمی‌کنند. همچنین در برخی زبانها مثل آلمانی، شروع کلمه با حرف بزرگ، برای جداکردن انواع کلمات نیست، بلکه همهٔ اسامی را با حرف اول بزرگ می‌نویسند.

۳-۶. زایش زبان طبیعی: تبدیل اطلاعات از پایگاه‌های دادهٔ رایانه‌ای به زبان قابل خوانش انسانی.

۳-۷.فهم زبان طبیعی: تبدیل بخش‌هایی از متن به بیانی که بیشتر فرمال باشد مثلا در ساختار منطق متغیرها تا رایانه آسان‌تر بتواند از آن استفاده کند.

۳-۸.نویسه‌خوانی نوری[1]: بازشناسی[2] خودکار متون موجود در تصاویر اسناد و تبدیل آنها به متون قابل جستجو و ویرایش توسط رایانه.

۳-۹.برچسب‌گذاری اجزای کلام: تعیین اینکه هر کلمه‌ای در جمله یا بخشی از متن چه نقشی دارد. این برچسب‌گذاری براساس نقش آن کلمه در متن، مانند اسم، فعل، قید، صفت، و غیره صورت می‌گیرد. بعضی کلمات ممکن است یک یا چند برچسب داشته‌باشند. اگر یک کلمه بیش از یک برچسب داشته‌باشد، نیاز به ابهام‌زدایی دارد. به عنوان مثال، book در انگلیسی ممکن است اسم یا فعل باشد. یا کلمه out حداقل پنج نقش مختلف می‌تواند بگیرد.

۳-۱۰.تحلیل نحوی: تعیین نمودار درختی تجزیه (تحلیل نحوی) جمله. این نمودار، نمودار درختی بنیادی و منظمی است که ساختار نحوی یک زنجیره (رشته) را مطابق با دستور زبان (گرامر) بافرض مستقل از متن ارائه می‌کند. نحو زبان طبیعی مبهم است و هر جمله‌ای تحلیل‌های ممکن متکثری دارد.

۳-۱۱.پاسخ‌گویی به پرسش‌ها: ارائهٔ پاسخ به پرسش‌هایی با زبان‌های انسانی. برخی پرسش‌ها پاسخ درست و مشخصی دارند. مثلاً «حرف اول کانادا چیست؟»، اما برخی پرسش‌ها پاسخ روشن و مشخصی ندارند، مثلاً «معنای زندگی چیست؟». در کارهای جدید به این‌گونه پرسش‌ها و دیگر پرسش‌های پیچیده هم توجه می‌شود.

۳-۱۲.استخراج روابط: تعیین نسبت‌ها و ارتباط‌ها میان هستارهای اسمی. (مثلاً اینکه چه کسی با چه کسی ازدواج کرد).

---

[1] . OCR
[2] . recognition

www.SID.ir

۳-۱۳.تعیین پایان جمله: یافتن حد پایانی جملات، یعنی جایی که هر جمله پایان می‌پذیرد. انتهای جمله‌ها معمولاً با نقطه یا دیگر علائم سجاوندی مشخص می‌شود، اما گاهی این علائم برای منظورهای دیگری استفاده می‌شوند (مثلاً نقطه برای نشان دادن اختصارات نیز به کار می‌رود).

۳-۱۴.تحلیل احساس: استخراج اطلاعات مربوط به گوینده سخنان بر اساس مجموعه‌ای از مدارک. از این اطلاعات برای تعیین تمایل نسبت به چیزی، و به‌ویژه برای امور بازاریابی و تبلیغات استفاده می‌شود.

۳-۱۵.تشخیص گفتار: هدف از تشخیص گفتار که در متون علمی بیشتر با نام بازشناسی گفتار شناخته شده‌است، طراحی و پیاده‌سازی سیستمی است که اطلاعات گفتاری را دریافت و متن و فرمان گوینده را استخراج می‌کند. فناوری بازشناسی گفتار به رایانه‌ای که توانایی دریافت صدا را دارد (برای مثال به یک میکروفن مجهز است) این قابلیت را می‌دهد که گفتار کاربر را متوجه شود. این فناوری در تبدیل گفتار به متن و یا به عنوان جایگزینی برای صفحه کلید یا ماوس برای وارد کردن دستورات مورد استفاده قرار می‌گیرد. سیستم‌های واکافت کننده گفتار انواع مختلفی دارند، بعضی قادرند گفتار پیوسته را شناسایی نمایند، بعضی دیگر فقط می‌توانند گفتار گسسته (که بین کلمات سکوت وجود دارد) را شناسایی کنند. همچنین سیستم‌ها قادرند واژگان گفته شده توسط افراد مختلف و یا فقط توسط یک گوینده تشخیص دهند. بهر حال ایده‌آل‌ترین سیستم آن است که بتواند گفتار پیوسته غیر وابسته به گوینده را در محیط نویزی شناسایی نماید. این سیستم‌ها با بکار گیری روش‌های مختلف طبقه‌بندی و شناسایی الگو قادرند به تشخیص واژگان هستند که البته برای افزایش دقت در شناسایی از یک فرهنگ لغات نیز در انتهای سیستم استفاده می‌شود.

۳-۱۶.تقطیع گفتار: عبارت از این است که رایانه بتواند پس از دریافت صدای شخص، آن را به کلمات تفکیک کند. این عمل، وظیفه‌ای فرعی ذیل تشخیص گفتار است.

۳-۱۷.جداسازی مباحث: جداسازی بخش‌هایی از متن که هر یک به مبحث متفاوتی اختصاص دارند.

۳-۱۸.جداسازی کلمات: جداکردن کلمات یک متن از یکدیگر. در زبان انگلیسی، کلمه‌ها با فاصله از یکدیگر جدا شده‌اند اما در برخی از زبان‌ها چنین نیست مثلاً در زبان‌های چینی و ژاپنی. جداسازی کلمات در این‌گونه زبان‌ها نیازمند دانش لغت‌شناسی و علم صرف کلمات در زبان مربوطه است.

۳-۱۹.رفع ابهام از معنای کلمات: بسیاری از کلمات بیش از یک معنا دارند. لازم است مشخص کنیم که کدام معنا در یک بافت خاص بیشتر از دیگر معانی مناسب است.

۴. پژوهش موسیقی موازی با پردازش زبان طبیعی

در ابتدا باید به بررسی این پرسش پرداخت که آیا تکنیک‌های تجزیه‌کردن در NLP را می‌توان در تجزیهٔ موسیقایی مورد استفاده قرار داد یا خیر. مشکل اصلی در موسیقی، همانند NLP، ابهام است: چندین ساختار مختلف ممکن است با یک سکانس موسیقایی همساز باشند، در حالی که شنونده معمولاً فقط یک ساختار را می‌شنود. بهترین تجزیه‌کنندهٔ احتمالاتی ما می‌تواند به‌درستی ۸۵,۹ درصد از عبارات را در آزمون مجموعه‌ای از هزار ترانهٔ محلی پیش‌بینی کن (Bod, 2001, p.35). برای درنظرگرفتن موسیقی به‌عنوان زبانی طبیعی، باید نشان داد که پردازش موسیقی در همان طبقه مسائل NLP قرار دارد. سطوح

www.SID.ir

پردازش متن توسط NLP، که در (جدول ۲) فهرست شده، از ضبط کردن (صدا، یا گفتار) تا فهم (معنای یک گفتار) را شامل می‌شود. این سطوح در مورد موسیقی نیز وجود دارد. موسیقی را شبیه به زبان طبیعی، می‌توان ضبط کرد و به‌عنوان یک شکل موج نشان داد. در سطح «آواشناسی» تلاش می‌شود که ساختار یک صدا، بررسی شود و نت‌ها یا سازها از یکدیگر تفکیک شوند. با وجود این، موسیقی در این عرصه بسیار پیچیده‌تر است و وظیفهٔ بازشناسی صدا با مشکلات اساسی مواجه است.

عرصه‌های پژوهش موسیقی	سطح NLP
تحلیل شکل موج، سیگنال‌های شنیداری	آواشناسی
شناسایی رویدادهای صوتی	واج‌شناسی
نمادهای متن موسیقی، نمادسازی	ریخت‌شناسی
مدل N-grams، تجزیه	نحو
هارمونی‌ها، سطح عبارت	معنی‌شناسی
برجسته‌سازی عبارت‌ها، صوت	کاربردشناسی
تفسیرها، بافت قطعه	گفتمان

جدول ۲. سطوح NLP با وظایف مربوطه در پژوهش موسیقی

دومین شباهت بسیار مهم از این امر حاصل می‌شود که در هر دو حوزه از نت‌نویسی نمادین استفاده می‌شود. متن‌های موسیقایی نیز از نویسه‌هایی تشکیل می‌شوند که نت خوانده می‌شوند. شبیه به ریخت‌شناسی و نحو NLP، موسیقی نیز ساختار دستورزبان- مانند پنهان و قواعدی پنهان دارد. بخشی از آن، هارمونی است. این قواعد مشخص می کند که چگونه کلمات (نت‌ها) کنار یکدیگر قرار بگیرند، و چگونه می‌توان با استفاده از آنها عبارات خوش‌فرم ساخت. نت‌ها و وابستگی آنها را می‌توان به نحو موسیقی مرتبط دانست اما پیشروی‌های هارمونیک را به کاربردشناسی قطعه. بالاترین سطح NLP (گفتمان) نیز در موسیقی به شکل ایده‌ها، میل یا اشتیاق (موسیقی رومانتیک) سازندهٔ اثر، و نیز تصاویر و کارهایی در پشت آن (موسیقی برنامه‌ای)، رواج دارد.

۵. تاریخچهٔ پردازش زبان طبیعی و پژوهش موسیقی

تاریخچهٔ NLP به ابتدای تاریخ محاسبه بازمی‌گردد، زیرا این باور وجود داشت که توانایی رایانه‌ها برای پردازش زبان طبیعی به همان صورتی که ما خود آن را پردازش می‌کنیم، محصول برجستهٔ ماشین‌های هوشمند واقعی خواهد بود. از این فرض چنین برمی‌آید که استفاده از زبان، جزء ذاتی استعداد شناختی انسان است. بر همین اساس، آزمونی معروف به آزمون تورینگ برای بررسی هوشمندبودن یک ماشین پیشنهاد شد. هدف اصلی از این آزمون این است که یک ماشین حقیقتاً هوشمند بتواند بحثی را با یک انسان به‌شکلی پیش ببرد که او نتواند تشخیص دهد که با یک انسان حرف می‌زند یا با یک ماشین. این تعریف از هوشمندی، پژوهش در NLP را با هدف نهاییِ آفرینش چنین برنامه‌ای برای مکالمه موجب شد. برای ممکن ساختن چنین هدفی، باید شیوه‌ای که مردم در تفکر، استدلال‌ورزی و صورتبندیِ اندیشه‌های خود پیش می‌گیرند، به‌طور بالفعل و فیزیکی پیاده‌سازی می‌شد.در همین مورد می‌توان شباهتی را میان موسیقی و گفتار آدمی مشاهده کرد. با فرض اینکه موسیقی زبانی طبیعی است، آزمون تورینگ می‌تواند شامل تولید قطعات موسیقی‌ای باشد که یک فرد خبره نتواند تشخیص دهد که یک انسان آنها را ساخته یا ماشین. اما چرا یک فرد خبره، و نه یک فرد عامی؟ زیرا برای آزمون تورینگ فرقی ندارد که طرف

www.SID.ir

صحبتش زبان گفتگو را بداند یا خیر. ما این آزمون را آزمون ملایم تورینگ می‌نامیم که برخلاف زبان‌های طبیعی قاعده‌مند انسانی، با موسیقی سازگارتر است.

۶.معرفی دستورزبان‌ها

در هر دو حیطهٔ NLP و پژوهش موسیقی، آثار پیشگامی وجود دارد. برخی تحلیل‌های زبان طبیعی به شکل نظریات زبانشناختی پیش از ورود ماشین‌های محاسبه‌گر ساخته‌شدند. در همان زمان، در حوزهٔ پژوهش موسیقی پیشارایانه‌ای می‌توان به آثار هاینریش شنکر و نظریهٔ فروکاست او در آغاز قرن بیستم اشاره کرد.این چیزها پس از اختراع رایانه به‌سرعت برای NLP تغییر کرد. مواجهه با رایانه به‌صورت متنی است و این قالب طبیعیِ فایل‌های رایانه‌ای است. افزون بر این، نیاز بسیار شدیدی به تحول ماشین‌های ترجمهٔ خودکار وجود داشت. آغاز این تحول، یعنی نظریهٔ نوام چامسکی دربارهٔ بافت دستورزبان‌های آزاد برای متون طبیعی، به سال ۱۹۵۶ باز می‌گردد. بازنمود و پردازش موسیقی اولویت اصلی آن زمان نبود. به‌عنوان کاری مشابه در حیطهٔ موسیقی می‌توان به کتاب «نظریهٔ زایشی موسیقی تونال» نوشتهٔ لرداف و جکندال[1] در سال ۱۹۸۳ اشاره کرد. هر دوی این رویکردها به شیوهٔ واحدی به این حیطه‌های مرتبط می‌پردازند، یعنی با معرفی دستور زبانی صوری که می‌تواند نمونه‌هایی را در این حیطه‌های مفروض، تولید کند(Lerdahl & Jackendoff, 1983).

۷. دوران استفاده

نخستین پژوهشگران NLP بسیار خوش‌بین بودند. این پژوهش‌ها با هدف تحول و گسترش ماشین ترجمهٔ خودکار پیش رفت. سیستم های مختلفی ساخته شد که گرچه آنها در مورد نمونه‌های بسیاری محدودی کارمی‌کردند اما در در کاربرد عالم واقع ناموفق بودند. این پژوهش به جایی رسید که دیگر نمی‌شد به چیز بیشتری دست یافت، در حالیکه آنها هنوز سیستمی قوی که بتواند بر روی داده‌های واقعی کار کند، نساخته بودند. این امر، کل این حیطه را زیر سوال برد. به نظر می‌رسید که باوجود آن سیستم‌های پیچیده، تقلید از شناخت آدمی در حیطهٔ زبان طبیعی ممکن نیست.در همان زمان بود که پژوهش موسیقی پا به میدان گذاشته بود. موسیقی به اندازهٔ زبان طبیعی با مشکل مواجه نبود زیرا بوسیلهٔ سیستمی که از موادی که به آنها می‌پرداخت فهم کاملی نداشت، فریب شنوندگان بی‌تجربه و در نتیجه، گذشتن از آزمون ملایم تورینگ ممکن بود(Bod, 2012).

۸.پردازش زبان طبیعی در حال حاضر

رویکرد قدیمی، یعنی ساختن مدلی که همهٔ مشکلات ما را در حیطهٔ شناخت گفتار آدمی حل کند، به نظر نادرست می‌رسد. پژوهش‌های اخیر NLP از یک سو بر ساختن مدل‌های زایشی دقیق‌تری که دربردارندهٔ احتمال یا دیگر جنبه‌های مشخصه‌های دیگر زبان‌ها باشند مبتنی است (مثلاً استفاده از ماتریس‌های ارزش مشخصه AVMs) . تکنیک‌های NLP در بازیابی اطلاعات نیز وارد شده‌اند تا اجرای این سیستم‌ها را ارتقاء بخشند. نهایتا اینکه، بهبود زیادی در ترجمهٔ ماشینی صورت گرفته، یعنی در هدف نخستینی که پژوهشگران NLP داشتند. مترجم گوگل[2] یا وِیوز روزی[3] نمونه‌های این امر هستند. پژوهش موسیقی کنونی احتمالاً همین مسیر را دنبال خواهد کرد.بررسی‌ها نشان می‌دهد که مدل‌های تجزیه احتمالاتی NLP را می‌توان به‌طور

[1] . Lerdahl & Jackendoff
[2] . Google translate
[3]. Wave's Rosy

Archive of SID
www.SID.ir

موفقیت‌آمیزی برای تجزیه موسیقایی استفاده کرد. بهترین تجزیه‌کنندهٔ احتمالاتی می‌تواند به‌درستی ۸۵،۹ درصد از عبارات را در آزمون مجموعه‌ای از هزار ترانهٔ محلی پیش‌بینی کند. که نتایج این بررسی ها می‌تواند مقدمه‌ای باشد برای دیگر مدلهای موسیقایی. تجزیه‌کنندهٔ ای برای بالابردن سرعت کار وقت‌گیر حاشیه‌نویسی بر ترانه‌های محلی نیز می‌توان به کار برد و در نتیجه برای خلق پایگاه‌های داده موسیقایی بزرگتر در موسیقی‌شناسی به کمک رایانه. (Bod, 2012, p. 7).

ارزیابی نتایج یک پژوهش نشان می‌دهد که طبقه‌ای از الگوهای موسیقایی وجود دارد، که عبارات پرشی خوانده می‌شود، که اصول گشتالتی نزدیکی و شباهت را به چالش می‌کشند. در این پژوهش سایت‌های مربوط به صدا و موسیقی را می‌توان بر حسب حضور یا عدم حضور محتوای موسیقایی و بر حسب حضور یا عدم حضور جامعهٔ کاربری برای ایجاد یا ویرایش اطلاعات بافتی، دسته‌بندی کرد.

سایت‌های نمونه	محتوای موسیقایی	جامعهٔ کاربری
صفحات وب هنرمندان، نشریات	خیر	خیر
آرشیوهای اینترنتی، آی تونز، اسپاتیفای	بله	خیر
فیس‌بوک، توئیتر، موزیک‌برینز، لست. اف‌ام	خیر	بله
فری‌ساند، ساندکلود	بله	بله

جدول۳. نتایج پژوهش مربوط به صدا و موسیقی (برحسب محتوای موسیقایی و جامعه کاربری)

۹.پژوهش‌های مرتبط

پژوهش‌های اخیر در موسیقی بر بازیابی اطلاعاتی موسیقی متمرکز شده است، هم در مورد بازنمودهای سیگنال و هم در مورد بازنمودهای موسیقی نمادین. در بیشتر موارد این پژوهش‌ها به مسائل اساسی‌ای می‌پردازند در خصوص اینکه چگونه رایانه‌ها باید به‌طور کلی با داده‌های موسیقایی برخورد کنند. سطح تفسیر موسیقایی وارد سطح معناشناسی نمی‌شود و دلیل آن احتمالاً این است که معنای موسیقی مبهم است. با وجود این، باید توجه داشت که بازیابی اطلاعاتی متنی درحال حاضر، از لایهٔ معناشناختی متن (طبقه‌بندی متن، وجودشناسی‌ها، و نسبت‌های میان اصطلاحات، وابستگی‌های متون، لایهٔ زبان‌شناختی متن) بهره می‌برد.شایسته است تا بر کار لردهال و جکنداف تأکید کنیم، که نخستین بار آنها رویکردی زایشی را توصیف کردند که می‌توان از آن در موسیقی استفاده کرد. آنها این رویکرد را به‌شیوهٔ زبانشناسی محاسباتی مطرح کردند و با استفاده از رویکرد قواعد ترجیح، خاطر نشان کردند که پیاده‌کردن قواعدشان در یک سیستم واقعی، امکان‌پذیر است. برای پیاده‌سازی سیستم آنها، باید بسیار منتظر می‌ماندیم زیرا جاافتادگی‌های آوایی مختلفی را باید تعریف می‌کردند و نیز تصورات اساسی مهمی را باید تعریف می‌کردند که برای انسانها قابل فهم بود اما پیاده‌سازی آن برای ماشین کار سختی بود. یکی از تلاش‌های اخیر که ATTA نام دارد، بوسیلهٔ تعریف چندین محدودیت مهم برای سیستم از پس مسائل پیاده‌سازی برآمده است. این پیاده‌سازی از سطح معناشناختی فراتر نمی‌رود و مسائل مربوط به هارمونی را پشت سر نمی‌گذارد (Masatoshi,2006).دیگر رویکرد زایشیِ قواعد ترجیحی، که عناصر بسیار مهمی از NLP و احتمال مدرن را معرفی می‌کند، را تمپرلی[1] مطرح کرد(Temperley, 2007). احتمالات و آمار، پیکره و بخش تفکیک‌ناپذیر همه نظریات NLP هستند، از این جهت که احتمالات می‌توانند معنای متن را با ارجاع دادن به وابستگی‌های درون آن شکل دهند (Manning, 1999). این کار، یادآور ایدهٔ

[1] . Temperley

www.SID.ir

دستورزبان‌های احتمالاتی است که قبلاً در مورد متن مطرح شده بود و بود[1] آن را در مورد موسیقی پیشنهاد کرد (Bod, 2002).تحلیل‌های آماری، عنصر بسیار مهمی در مدل‌های NLP هستند و نقشی اساسی در پژوهش موسیقی ایفا کرده و می‌کنند (Cope, 1992) در بسیاری از موارد، پاسخ بسیاری از مسائل که نتایج خوبی در متن‌ها داشته‌اند، در حیطهٔ موسیقی نیز می‌تواند به نتایج خوبی بیانجامد. به عنوان مثال، روش n-gram در مورد مشخصهٔ مؤلف که برای متون زبان طبیعی شکل گرفت، در مورد بازشناسی مصنف قطعهٔ موسیقی نیز نتایج خوبی داشت(Keselj, 2003).

۱۰.آیندهٔ پژوهش موسیقی

بنا به این فرضیه که NLP و موسیقی محاسباتی کنونی، به معنایی واحدند اما در دو حیطهٔ شبیه، و نه همانند، اجرا می‌شوند، هر دو حیطه می‌توانند از این میراث بهره‌مند شوند. برای نمونه، برنامه‌های کاربردی‌ای که سطوح زیادی از NLP را در بر می‌گیرند، اگر فقط بر سطوح خاصی تمرکز کنند بهتر کار خواهند کرد. همان‌گونه که اشاره کردیم، برخی از لایه‌های NLP به خوبی بر مادهٔ موسیقایی منطبق نمی‌شود. بیشترین کار انجام شده، در حیطهٔ آواشناسی، واج‌شناسی و ریخت‌شناسی موسیقی صورت گرفته است. اخیراً کارهایی در حیطهٔ معناشناسی نیز صورت‌گرفته اما در سطوح بسیار جالب‌تر و البته دشوارتر بالایی، یعنی معناشناسی و کاربردشناسی، گفتمان، مدلی مطرح نشده است. این حیطه‌ها معنای داده‌ها را مشخص می‌کنند و به فهم ساختار قطعه و جریان یافتن اندیشه مصنفان در آن می‌پردازد. برخی از وظایفی که به پژوهش موسیقی مربوط است و NLP به‌خوبی آنها را توسعه‌بخشیده، عبارتند از تحلیل عاطفی، طبقه‌بندی ژانر، خلاصه‌سازی خودکار و استخراج سبک. اگرچه هنوز روشن نیست که چگونه باید معنای موسیقی را در قالب وظایف محاسباتی بازنمایی کرد، اما تکنیک‌های رویکرد آماری و استخراج داده می‌توانند ابزارهای مناسبی برای توصیف این پدیدار باشند.منابع شنیداری و موسیقایی‌ای که گروه‌های مختلف کاربران بر روی اینترنت می‌گذارند، به‌طور تصاعدی در حال افزایش است. محتوای رسانه‌ای اجتماعی را می‌توان در سطوح مختلف ساختاری مشاهده کرد، و کاربرانی هم که با آنها مواجهند، ممکن است حرفه ای یا غیرحرفه‌ای باشند. برداشت و سازمان‌دادن به این اطلاعات از نظر معناشناختی می‌تواند برای بافت‌شناسی در بازیابی اطلاعات موسیقایی مفید باشد. روش پیشنهادی ترکیبی است از تکنیک‌های کاوش رسانه اجتماعی، استخراج اطلاعات و پردازش زبان طبیعی، تا به این وسیله اطلاعات معنادار بافتی را از داده‌های اجتماعی استخراج کنیم. (Oramas, 2014, p. 20).همان‌گونه که Schedl بیان‌کرده، عوامل مؤثر بر ادراک انسان از موسیقی را می‌توان در قالب محتوای موسیقیایی، بافت موسیقایی، بافت کاربری، و خصوصیات کاربری دسته‌بندی کرد بر اساس این طبقه‌بندی، به نظر می‌رسد که بافت موسیقایی و بافت کاربری جنبه‌های کلیدی‌ای از بازیابی اطلاعات موسیقایی هستند و گروه‌های آنلاین، مکان مناسبی برای جستجو دربارهٔ این نوع اطلاعات می‌باشند (Schedl, 2013, pp. 205-213).

۱۱. نتیجه‌گیری

باوجود اهمیت و گستردگی موسیقی در سطح جوامع بشری، حوزهٔ پژوهش دربارهٔ بازیابی اطلاعات موسیقایی نسبتاً جدید است و کمتر از دو دهه از آغاز آن می‌گذرد. با این حال، این حوزه از همان ابتدا تاکنون به عنوان یک حوزهٔ پژوهشی، روند روبه جلویی را داشته است. برخی از مهم‌ترین دلایل موفقیت در این حوزه عبارتند از:

۱) توسعهٔ تکنیک‌های فشرده‌سازی موسیقی در اواخر دههٔ ۱۹۹۰

[1] . Bod

www.SID.ir

۲) افزایش توان محاسباتی رایانه‌های شخصی، که کاربران و اپلیکیشن‌ها را قادر ساخته تا مشخصه‌های موسیقایی را در زمان معقولی استخراج‌کنند.

۳) دسترسی گسترده به دستگاه‌های قابل‌حمل اجراکنندهٔ موسیقی

۴) شکل‌گیری سرویس‌های جاری‌سازی موسیقی همچون اسپاتیفای، گروشارک، آردیو و دیزر که نویدبخش استفادهٔ نامحدود از موسیقی در هرزمان و هر مکانی هستند. (Serra, 2013, p.210).

از آنجا که موسیقی از داده های دیجیتالی گسترده و بازنمودهای فراوانی برخوردار است، در سطوح معناشناختی متعددی جای می گیرد. به طور معمول، مجموعه‌های موسیقی دیجیتال دربردارندهٔ اسناد دیجیتالی متعددی هستند که مربوط به هر اثر موسیقی‌است (Müller, 2007, p. 17).حوزهٔ پژوهش موسیقی با پژوهش در NLP شباهت دارد. هر دو حیطه بر روی گونه‌های مشابهی از داده‌ها که از ویژگی‌های مشترکی برخوردارند، کار می‌کنند. هر دو حوزه با داده‌هایی کار می‌کنند که انسان‌ها به راحتی آنها را درک می‌کنند اما برای آنکه آنها را به‌طور کامل برای رایانه‌ها قابل فهم سازند با مشکلات فراوانی مواجهند. برای پژوهش در هر دو حوزه از تکنیک‌های مشابهی استفاده می‌شود، و باید در خصوص هر یک از حیطه‌هایی که در دیگری پیشرفت بیشتری حاصل شده از هم استفاده کنند. پژوهشگران موسیقی می‌توانند دستاوردهای خود را در وظایفی مانند جداسازی صداها یا کشف محدوده‌های اصوات عرضه کنند و در مقابل از NLP در خصوص مواردی همچون روش‌های آماری، رویکردهای خودکار نسبت به معناشناسی، کمک‌گرفتن از بازیابی اطلاعات و داده‌یابی از طریق فهم زبان طبیعی، استفاده‌کنند. ضرورتی ندارد که همهٔ این تکنیک‌ها و رویکردهای به‌ارث‌رسیده کار کنند، اما یقیناً ارزش تلاش‌کردن دارند.

مراجع:

Bod, Rens (2001). Memory-Based Models of Music Analysis: Evidence against the Gestalt Principles in Music. Proceedings International Computer Music Conference Havana, Cuba.

Bod, Rens (2002) . A unified model of structural organization in language and music. JAIR, 7(1): pp.289–308.

Bod, Rens (2012). Probabilistic Grammars for Music. ILLC: University of Amsterdam.

Brants, Thorsten (2015). Natural Language Processing in Information Retrieval

Byrd, Donald (2002) .Problems of Music Information Retrieval in the Real World, University of Massachusetts Amherst.

Cope, David (1992).Computer modeling of musical intelligence in emi. Comp. Music Journal. 16(2):pp. 69–83.

Kassler, M. (1970). MIR:A Simple Programming Language for Musical Information Retrieval. In Harry B. Lincoln (Ed.), The Computer and Music (pp. 299–327). Ithaca, New York: Cornell University Press.

Keselj, Vlado (2003). N-gram-based author profiles for authorship attribution. In Proc. of the PACLING03 Conference.

Lerdahl, Fred and Jackendoff, Ray (1983). A Generative Theory of Tonal Music. The MIT Press.

Mahedero, Jose P. G. and Mart, Alvaro (2005). Natural language processing of lyrics. Proceedings of the 13th ACM International Conference on Multimedia. Singapore, November 6-11.

Manning , Christopher D and Schutze, Hinrich (1999). Foundations of Statistical Natural Language Processing. The MIT Press.

www.SID.ir

Masatoshi Hamanaka, Keiji Hirata, and Satoshi Tojo (2006). Implementing a generative theory of tonal music. Journal of New Music Research, 35, pp.249–277.

Muller, Meinard (2007). Information Retrieval for Music and Motion. New York: Springer.

Nadkarni, P. M., Ohno-Machado, Lucila and Chapman, Wendy (2011). Natural language processing: an introduction. J Am Med Inform Assoc. 18(5): pp. 544–551.

Oramas, Sergio (2014). Harvesting and Structuring Social Data in Music Information Retrieval. Barcelona, Spain: Universitat Pompeu Fabra.

Schedl, M. and Schnitzer, D. (2014). Location-Aware Music Artist Recommendation. Lecture Notes in Computer Science, vol. 8326, pp. 205-213.

Schedl, M. (2013). On the Use of the Web and Social Media in Multimodal Music Information Retrieval. Postdoctoral Thesis (Habilitation).

Serra, X. (2013). Exploiting Domain Knowledge in Music Information Research. Stockholm Music Acoustics Conference 2013 and Sound and Music Computing Conference 2013.

Temperley, David (2007). Music and Probability. The MIT Press.

Wołkowicz, Jacek, Kulka, Zbigniew and Vlado Keselj (2008). N-gram-based approach to composer recognition. Archives of Acoustics, 33(2008)(1):43–55, Jan 2008.

Wołkowicz, Jacek and Valdo Keselj (2010) .Predicting Development of Research In Music Based On Parallels With Natural Language Processing, 11th International Society for Music Information Retrieval Conference.

Archive of SID

www.SID.ir

Natural Language processing and Music Information retrieval

Nahid Amiri

Department of Knowledge and information Science, Faculty of Psychology, Kharazmi University, Iran, nahidamirilib@gmail.com

Abstract: The Purpose of this paper is to study the tasks of Natural language processing (NLP) in Music information retrieval. Music computational research in the field suggests that the area of NLP is close to recent research in music. So we can profit from this similarity Search NLP techniques applied in music. In this paper, the similarity between both the above described area. NLP levels listed here refer to relevant tasks in computational study music. NLP brief introduction about the history of music in the background makes clear the following different aspects of the research about the music, given its proximity to the NLP, is introduced. Finally, recent research on Music computational research compared with similar ideas in NLP and presented guidelines for Music information retrieval.

Keywords: Information retrieval, Natural language processing, Music information retrieval.

Archive of SID

www.SID.ir