

Recunoașterea automată a acordurilor acustice

Linca Răzvan Cosmin

05-04-2020

1 Definirea

Subiectul propus urmărește prezentarea și implementarea unui proces complex obținut prin îmbinarea a două domenii aparent diferite: muzica și inteligența artificială.

Prima parte (recunoașterea automată a acordurilor acustice), reprezentând domeniul muzical, definește procesul automat capabil să analize o mostră muzicală, să determine succesiunea continuă de acorduri diferite și să reprezinte automat această succesiune într-un format standard, numit tabulatură muzicală (format specific acordurilor acustice). Analiza unei mostre muzicale și extragerea unor trăsături specifice se realizează aplicând algoritmi de procesare pentru semnalul audio.

Partea a doua (inteligența artificială), reprezintă unealta care stă la baza proceselor de învățare și deosebire a trăsăturilor unor acorduri din cadrul unei mostre muzicale. Se vor enunța și prezenta în detaliu câțiva algoritmi de învățare automată cu aplicabilitate pentru problema noastră, cu scopul de a ajunge treptat la un algoritm complex capabil să analizeze automat semnalul audio, să clasifice cu o precizie cât mai mare fiecare acord, fără a fi necesară intervenția umană în corectarea rezultatului.

2 Motivarea

Muzica acustică contemporană se află într-o perioadă de evoluție din punct de vedere tehnologic, odată cu evoluția paralelă a rețelelor de socializare (platforme video și de streaming, platforme de socializare ș.a.m.d), care au conectat artiști profesioniști și amatori ai domeniului.

Această evoluție a determinat o depărtare a interpretului de partitură, acesta fiind ori un artist profesionist care învață un cântec prin simpla ascultare sau vizualizare a unei înregistrări, sau un amator pasionat care învață prin urmărirea repetată a unor tutoriale aflate pe platformele online. De altfel, se spune că mulți compozitori și chitariști renumiți nu puteau nici măcar să citească partiturile muzicale, printre care se numără Jimi Hendrix, Eric Clapton, Elvis Presley sau interpreții trupei The Beatles. [3]

Chitariștii susțin că notația clasică prin partitură muzicală este foarte bine orientată către instrumente cu clape, cum ar fi pianul, fiind în general neutilizată pentru chitară. Pe de altă parte, notația prin tabulatură este optimizată pentru instrumentele cu corzi și freturi, fiind de folos mai mare pentru aceștia. Cu toate acestea, o cantitate uriașă de piese au fost scrise doar în notația clasică prin partitură, iar transcrierea unei piese de la partitură la o tabulatură corectă este un proces complex care necesită cunoștințe de specialitate. Prin urmare, materialul devine inaccesibil pentru chitariștii aspiranți care nu sunt atât de specializați.

Principala motivație a realizării proiectului are astfel la bază dorința de a automatiza procesul de transcriere a conținutului audio direct într-o reprezentare simplă și sugestivă, folosind notația prin tabulatură.

O motivație secundară are la bază provocarea lansată de îmbinarea dintre cele două domenii aparent diferite: muzica și inteligența artificială. Este o oportunitate pentru oricine abordează aceste subiecte de a acumula informații, abilități și deprinderi noi, prin analiza și studierea inițială a domeniilor, iar apoi prin propunerea și descrierea unei soluții.

3 Analiza problemei

Recunoașterea automată a acordurilor muzicale (RAAM) a fost un domeniu cercetat în mod activ în domeniul obținerii informațiilor muzicale, *Music Information Retrieval (MIR)*, în ultimii 20 de ani. Această categorie de algoritmi sunt o parte esențială a multor aplicații muzicale, cum ar fi sisteme de transcriere automată pentru diverse instrumente, aplicații de învățare în cadrul educației muzicale sau algoritmi de recomandare a muzicii sau a genurilor muzicale.

RAAM urmează, în general, o paradigmă de calcul precisă. Acestea implică două faze distincte:

- Faza 1: Extragerea unor descriptori specifici din semnalul audio. Această etapă poartă numele de *extragerea trăsăturilor (feature extraction)*. Extragerea și analiza ulterioară a trăsăturilor are la bază algoritmi de prelucrare și procesare a semnalului audio. Domeniul de procesare a semnalului oferă diferite reprezentări posibile ale unui semnal audio care pot fi utilizate pentru probleme de clasificare, cum ar fi RAAM. Considerăm, spre exemplu, utilizarea Transformatei Fourier în prelucrarea semnalului sonor. Procesul de extracție a caracteristicilor se încheie odată cu crearea unei reprezentări sugestive a semnalului. Reprezentarea care va fi utilizată în implementare este chromagrama.

Chromagrama oferă o descriere suficient de potrivită din punct de vedere muzical a semnalului audio, fiind utilizată, în acest caz, pentru a determina cel mai probabil acord care se cântă în acel interval de timp. Rezultatul final al procesului este determinat de o reprezentare de tipul note versus timp, (*pitch class versus time*).

- Faza 2: Clasificarea acordurilor pe baza reprezentării determinate în prima fază. Această etapă explorează diferite metode de clasificare cu aplicabilitate pentru problema RAAM.

Se vor enunța și analiza în detaliu mai multe metode specifice, pe baza unor cercetări conexe ale domeniului. Astfel, se vor descrie cronologic algoritmi potriviți, pornind de la modelele probabilistice, ca modelele Markov cu stări invizibile, *Hidden Markov Models (HMMs)*, fiind prima abordare a problemei enunțată în mai multe lucrări științifice, până la construirea și optimizarea unor rețele neuronale convoluționale, utilizate recent în acest domeniu.

4 Descrierea soluției propuse

Pentru construcția modelului de recunoaștere acustică, am ales utilizarea unei rețele neuronale convoluționale (CNN). CNN-urile au crescut în popularitate în ultimii ani, mai ales în domeniul viziunii computerizate (*computer vision*).

În cadrul rețelelor neuronale convoluționale, imaginea construită sau aleasă este folosită ca parametru de intrare pentru CNN, iar imaginea este trecută prin mai multe straturi, cum ar fi straturile convoluționale, straturile de unificare (*pooling layer*) și straturile de activare (*activation layer*). Straturile convoluționale sunt construite din mai multe filtre, care pot fi interpretate individual, fiecare învățând unele caracteristici superioare ale imaginii.

În problema de RAAM, detectarea acordurilor poate fi tratată în mod similar cu recunoașterea imaginilor, deoarece se pot crea imagini cu reprezentarea de tip chromagramă. Identificarea notelor sau a acordurilor este mai simplă decât clasificarea imaginilor, întrucât nu implică texturi importante de învățat, rotații sau scalări. [1] Cu toate acestea, RAAM vine cu alte provocări. Notele muzicale din cadrul chromagramei nu sunt localizate într-o singură regiune în același mod în care sunt cele mai multe obiecte din imagini - o notă la o anumită frecvență fundamentală va fi compusă din armonice la multipli acelei frecvențe.

În ciuda acestei provocări, CNN-urile au proprietăți avantajoase care pot fi aplicate în cadrul RAAM. Experimentele anterioare au sugerat că agregarea informațiilor considerând mai multe cadre muzicale din același sample, determină obținerea unei predicții cu o performanță mai mare. Astfel, convoluțiile determinate asupra datelor de intrare permit modelului creat să învețe caracteristici muzicale polifonice valoroase.

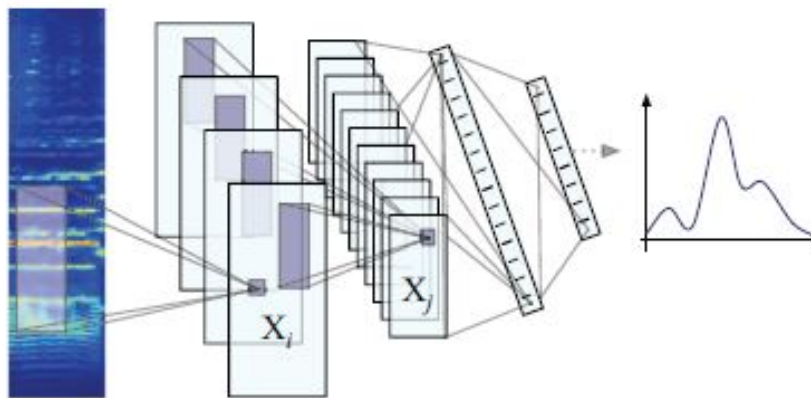


Figura 1: Reprezentare CNN [2]

Bibliografie

- [1] Michael Bereket and Karey Shi. “An AI Approach to Automatic Natural Music Transcription”. In: *Stanford University* (2017).
- [2] Eric J. Humphrey and Juan P. Bello. “Rethinking Automatic Chord Recognition with Convolutional Neural Networks”. In: *Music and Audio Research Lab (MARL) New York University* (2012 11th International Conference on Machine Learning and Applications).
- [3] *The Music Studio*. <https://www.themusicstudio.ca/blog/2017/11/909/>. Accessed on 2020-03-29.