



EÖTVÖS LORÁND TUDOMÁNYEGYETEM

INFORMATIKAI KAR

PROGRAMOZÁSELMÉLET ÉS SZOFTVERTECHNOLÓGIAI
TANSZÉK

Automatikus zenei hangszerfelismerés többszólamú zenében mély neuronhálók segítségével

Témavezető:

Gombos Gergő

Adjunktus, PhD

Szerző:

Hamrák János

programtervező informatikus MSc

Budapest, 2020

Az eredeti szakdolgozati / diplomamunka témabjelentő helye.

Tartalomjegyzék

1. Bevezetés	3
1.1. Motiváció	4
1.2. A dolgozat felépítése	5
1.3. Kapcsolódó munkák	5
2. Elméleti háttér	7
2.1. Zene és reprezentációi	7
2.1.1. Zene fogalma, tulajdonságai	7
2.1.2. Hang reprezentációk	7
2.1.3. Music Information Retrieval (MIR)	8
2.2. Mesterséges intelligencia	10
2.2.1. Gépi tanulás (machine learning)	11
2.2.2. Mély tanulás (deep learning)	11
3. Adathalmaz	12
3.1. Kiválasztási szempontok	12
3.2. Philharmonia Orchestra	12
3.3. OpenMIC	12
4. Módszertan	13
4.1. Előfeldolgozás	13
4.2. Architektúra	13
4.3. Megvalósítás	13
4.3.1. Könyvtárak	13
5. Kísérletek, eredmények	14
5.1. MÉRŐSZÁMOK	14

5.2. Eredmények	15
5.2.1. próba1	16
5.2.2. próba2	16
5.2.3. próba3	16
5.2.4. próba4	16
5.3. Összehasonlítás	16
5.3.1. saját próbálkozásaim	16
5.3.2. más munkákkal (hagyományos ML)	16
6. Összegzés, kitekintés	17
Irodalomjegyzék	18

1. fejezet

Bevezetés

Napjainkban a zenéhez legkönnyebben digitális formában, a világhálón keresztül férhetünk hozzá. Néhány kattintással olyan zenei tartalomszolgáltatókat érhetünk el, melyek széleskörű adatbázissal rendelkeznek. A folyamatosan bővülő adatmennyiség ellenére ezeknek az adatbázisoknak átláthatónak és könnyen kezelhetőnek kell maradniuk, hogy a felhasználókat a kívánt módon tudják kiszolgálni. Ennek érdekében nap mint nap új megoldások születnek zenei információk automatikus kinyerése és feldolgozása céljából. Ezek teszik lehetővé a digitálisan tárolt zenék körében például az osztályozást vagy keresést.

A zenei információk kinyerésének tudományába (Music Information Retrieval, a továbbiakban: "MIR") tartozik a továbbiakban taglalt probléma, az automatikus hangszerfelismerés. Ez egy osztályozási feladat. Célja, hogy a meglévő digitális hanganyag alapján az adott zenéről megállapítsuk, hogy milyen hangszerek szólalnak meg benne. Ezt az információt több célra is fel tudjuk használni, például:

- Későbbi feldolgozásra, további MIR feladatok inputjaként
- Statisztikák készítésére
- Adatbázisban való keresés szűrőfeltételeként
- Egy ajánlórendszer részeként, ahol az aktuális zeneszámot követően például egy hangszerelésében hasonló számot szeretnék ajánlani a felhasználónak.

Az automatikus hangszerfelismerés feladatot több aspektusból lehet megközelíteni, például a bemeneti adatok jellege, reprezentációja, a megvalósított architektúra,

az osztályozás módszere, vagy az osztályok száma alapján. A dolgozatom keretein belül felkutatok néhány létező megoldást, majd ezekre alapozva prezentálom saját megközelítésemet és ennek eredményeit. Az általam bemutatott megoldás egy multi-class multi-label osztályozást valósít meg mély neuronhálós rendszer segítségével többszólamú zenében.

1.1. Motiváció

Az ember kognitív képességei segítségével a zenében könnyedén fel tudja ismereni az egyes hangszereket. Ugyanez a feladat a számítógép számára azonban már sokkal kevésbé triviális. Ennek egyik oka, hogy egy hangszer megszólaltatásának digitális reprezentációja nagyon változatos lehet. Függ például a hangszíntől, hangmagasságtól, hangerőtől és előadásmódtól, de a felvétel minőségétől és az esetleges háttérzajtól is. További nehezítő körülmény a többszólamúság, amikor egy időben több hangszert is megszólaltatunk, ezzel összemosva az egyszólamú környezetben is sokváltozós képünket.

A MIR nagyban támaszkodik a mesterséges intelligenciára. A számítógépek számítási kapacitásának folyamatos növekedése és az elérhető adathalmazok gyarapodása által pedig egyre nagyobb figyelmet kap a mesterséges intelligencia egy kiemelten számításigényes részterülete: a mély tanulás. Ezt bizonyítja, hogy az évente megrendezésre kerülő ISMIR (International Society for Music Information Retrieval) konferencián 2010-ben még csak 2 ([1], [2]) mély tanulással kapcsolatos cikk jelent meg, de 2015-ben már 6, 2016-ban pedig már 16. [3]

A mély tanulás tehát egy ígéretes módszer lehet a MIR problémák megoldásában, ideértve az automatikus hangszerfelismerést is. Ezt kihasználva és megoldás életszerűségére törekedve döntöttem úgy, hogy elkezdek kísérletezni mély neuronhálókkal többszólamú zenében. Célom volt találni egy tanításra alkalmas adathalmazt, azon pedig tervezni egy olyan mély neuronhálós rendszert, amely a jelenlegi megoldások pontosságát meghaladja.

1.2. A dolgozat felépítése

Dolgozatomban tehát az eddigi kapcsolódó kutatásokat, illetve saját munkám eredményét dolgozom fel. A következő alfejezetben felsorolom az általam relevánsnak tartott, a State-of-the-Arthoz vezető kutatásokat.

A második fejezetben betekintést adok a téma elméleti hátterébe. Először kifejtem a zenével kapcsolatos főbb fogalmakat, bemutatom fontosabb tulajdonságait, reprezentációit. Kitérek a MIR bemutatására is. Ezután bevezetem a gépi tanulás és a mély tanulás fogalmát.

A harmadik fejezet az adathalmazokról fog szólni. Itt előbb felsorolom az adathalmazok kiválasztásának szempontjait, majd minden felhasznált adathalmaznak ismertetem a főbb jellemzőit.

A negyedik fejezetben a módszertanról ejtek szót. Itt kifejtésre kerülnek az adatok előfeldolgozási módszerei, az általam bemutott mély tanulási architektúrák, illetve ezek megvalósításai.

Az ötödik fejezetben részletezem az általam végzett kísérleteket és ezek eredményeit. Ezeket összevetem egymással, illetve a releváns State-of-the-Art kutatásokkal.

A hatodik fejezetben összegzem a leírtakat, valamint továbbgondolom a kutatásomat, felvázolok néhány ötletet annak jövőjéről.

1.3. Kapcsolódó munkák

Az automatikus hangszerfelismerés témában a korábbi kutatások túlnyomó része a monofónikus, azaz egyhangszeres zenékkal foglalkozik. Martin és Kim [4] mintafelismerési statisztikai technikája 1023 izolált hangjegy és 15 különböző hangszer között a hangszercsaládok felismerésében 90%-os, egyéni hangszerek felismerésében pedig 70%-os pontosságot produkált. Brown [5] a kepsztrális együtthatókat használta fel K-közép klaszterezési módszeréhez. Eronen és Klapuri [6] széleskörű, spektrális és időbeli feature-halmaz segítségével - összesen 43 különböző feature felhasználásával - 81%-os hangszer és 95%-os hangszercsalád pontosságról számolt be. Deng [7] klasszikus zenei hangszerek tekintetében elemezte a különböző, gépi tanulási módszerekben használatos feature összeállításokat. Bhalke [8] tört Fourier-transzformáción alapuló MFCC feature-ök segítségével tervezett CPNN osztályozót mutatott be, amellyel

hangszer családok tekintetében 96.5%-os, hangszerek tekintetében pedig 91.84%-os pontosságot ért el.

Többszólamú környezetbe való átültetéssel foglalkozott Burred tanulmánya [9], aki a többszólamúságot két kísérlettel közelítette meg. Először csak egy-egy hangjegyet kombinált össze többszólamú hangjeggyé. Itt két szimultán hangjegy esetén 73.15%-os, három hangjegyre 55.56%-os, négy hangjegy kombinációjára pedig 55.18%-os pontosságot sikerült elérni. Másik kísérletként hosszabb szekvenciákat kombináltak össze, ekkor két hang esetén 63.68%-os, három hang esetén pedig 56.40%-os pontosságot kaptak.

Eggink és Brown [10] a polifónikus zenékben a hiányzó adat elméletükkel próbálták feltárni az egyes hangszereket. Ennek lényege, hogy felderítették azon idő- és frekvenciabeli részeket a zenén belül, ahol szeparáltan egy hangszer tulajdonságait vélték felfedezni és ezt dolgozták fel. Erre a módszerre épített Giannoulis és Klapuri [11] kutatása is, és hasonló megközelítést alkalmazott Garcia [12] is.

Jiang [13] egy többlépcsős megoldást mutat be. Első lépésben a hangszercsaládot határozták meg, ezzel szűkítve a lehetséges hangszerek halmazát és a változók számát. A pontos hangszer-meghatározás csak ezután következett.

Az előbbi kutatások többnyire hagyományos gépi tanulási megoldásokat alkalmaztak, amelyekhez maguk nyerték ki a különböző bemeneti feature-öket. Humphrey [14] írásában a mély tanulási architektúrákat ismerteti a MIR terület korszerű irányzataként. A témában gyakorlati segítségként szolgál Choi [15] írása, amiben konkrét adatrepresentációkat, mély neuronhálós rétegeket, és mély tanulási technikákat mutat be.

Li [16] a nyers hanganyagot inputként felhasználva egy konvolúciós mély neuronhálós rendszert mutatott be a polifónikus zenében való automatikus hangszerfelismerés kapcsán. Ezt a megoldást aztán összevetette hagyományos gépi tanulási módszerekkel is. A mély neuronhálós rendszer teljesített legjobban. 75.60%-os pontossággal, 68.88%-os felidézéssel, 72.08 mikro F értékkel és 64.33 makro F értékkel. Han [17] szintén egy mély konvolúciós hálót használt, azonban az osztályozás szempontjából máshogyan járt el: a zenékben egy darab domináns hangszert keresett. Bemenetként a zenék spektogramját használta fel, 0.602-es mikro és 0.503-as makro F értéket ért el.

2. fejezet

Elméleti háttér

Ebben a fejezetben a dolgozathoz kapcsolódó fogalmakat és elméleti alapokat mutatom be. Először magának a zenének a releváns tulajdonságairól ejtek szót. Ezután ismertetem a MIR kutatási területet, amelybe dolgozatom is tartozik. Majd végül a mesterséges intelligencián alapuló megoldásokról nyújtok elméleti bevezetőt, érintve a hagyományos gépi tanulás és a mély tanulás módszereit is.

2.1. Zene és reprezentációi

2.1.1. Zene fogalma, tulajdonságai

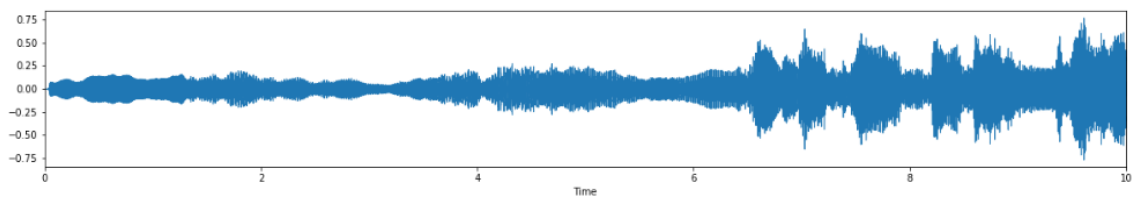
A zene egy meglehetősen összetett fogalom. Az ember számára a zene megjelenhet hang formájában, leírhatjuk őket szimbólumok segítségével egy kottában, előfordulhat szöveges formában dalszöveggként, képi formában egy albumborító, vagy egy zenész képében, illetve mozdulatokban egy zenei előadás keretében.

//TODO Often music means the audio content, although otherwise its scope extends to other types of musical information e.g., lyrics, music metadata, or user listening history.

2.1.2. Hang reprezentációk

A hangok fizikai mivoltukban rezgéseként jelennek meg. A rezgéseket matematikailag olyan folytonos függvényekkel tudjuk leírni, melyek értelmezési tartománya az idő, értékkészlete pedig a nyugalmi állapothoz viszonyított pillanatnyi kitérés.

Ilyen lehet például egy szinuszgörbe. Ahhoz, hogy a hangokat számítógépen tudjuk tárolni és feldolgozni, ezeket a függvényeket kell ábrázolnunk. Mivel azonban a számítógép számábrázolása véges, ezért a hangokat először digitalizálni kell. Ez azt jelenti, hogy a folytonos függvényeket diszkrét, azaz véges helyen vett és véges értékekkel rendelkező függvényekké alakítjuk. Ez úgy történik, hogy az eredeti függvényünkéből megadott időközönként mintát veszünk, diszkrét értékre kerekítjük, és ezeket az értékeket összefűzzük. Az így kapott függvény lesz a hang digitális reprezentációja.



2.1. ábra. Tíz másodperces hanganyag digitális reprezentációja

A hangok digitális reprezentációja tehát lényegében egydimenziós, mivel egy függvénygörbének tekinthetjük. Ezt szokták hívni nyers hangnak is, ugyanis további reprezentációkká tudjuk transzformálni. A MIR területen megjelenő mély tanulási megoldások jelentős része ezen nyers hangábrázolás helyett inkább a kétdimenziós reprezentációkat alkalmazza bemenetként. Ezt azzal indokolják, hogy a nyers bemeneten való tanítás sikeréhez nagyobb adathalmaz szükséges, mint a kétdimenziós reprezentációkéhoz. [15]

//TODO 2d reprezentációk

2.1.3. Music Information Retrieval (MIR)

A bevezetőben már említettem a zenei információk kinyerését (music information retrieval - MIR). Ez egy interdiszciplináris kutatási terület, magában hordozza többek között a zeneelmélet, pszichoakusztika, pszichológia, informatika, jelfeldolgozás és gépi tanulás tudományágakat. Céljára jól utal az elnevezése, zenéből szeretnénk releváns információt kinyerni, és ezeket felhasználni [15]. A felhasználásra szerintem nagyon jó, életszerű példát ad Downie 2003-as cikkének [18] bevezetője, amelyet a következőképp fordíthatunk le:

”Képzeljünk el egy világot, ahol egyszerűen felénekelhetjük egy számítógépnek a dalrészletet, ami már reggeli óta a fejünkben jár. A gép elfogadja a hamis énekünket, kijavítja, és azonnal javaslatot tesz arra, hogy éppen a ”Camptown Races” című számra gondoltunk. Miután mi belehallgatunk a gép által talált számos relevánsnak tartott MP3 fájl egyikébe, elfogadhatjuk a gép javaslatát. Ezután elégedetten elutasíthatjuk a felajánlást, hogy az összes további létező verzióját is felkutassa a dalnak, ide értve a nemrég megjelent olasz rap verziót, vagy a skótdudás duettre írt kottát.” [18]

Figyeljük meg, hogy ez a hétköznapi eset mennyire összetett probléma. A következő feladatok jelennek meg:

- Az emberi éneklés, vagy dúdolás alapján hangfelismerés.
- Hang alapú lekérdezés egy zenei adatbázisban az előbbi bemenettel.
- Hangelemzés, feldolgozás, hogy a hamis hangokat ki tudjuk javítani, az esetleges háttérzajokat eltávolítsuk, illetve ha kell, a dallamból automatikusan kottát generáljunk.
- Hasonlóságon alapuló keresés zenék között, hogy megtaláljuk a kívánt dalt az adatbázisban.
- Zenei feldolgozások detektálása, hogy további verzióit is megtaláljuk egy adott dalnak.

MIR problémák definiálását több szempontból közelíthetjük meg. Choi cikke [15] két tengelyre osztja fel a problémateret: szubjektivitás és eldöntési időmérték. A szubjektivitás tengelyen léteznek szubjektívebb feladatok, melyekre nincsenek egyértelmű válaszok. Ilyen lehet például a zene műfajának meghatározása. Objektívebb feladatoknak tekinthetjük azokat, melyek eredménye egyértelműen meghatározható, esetleg számszerűsíthető. Ide tartozik a hangszerfelismerés, vagy a tempó észlelés. [15]

A másik tengely, az eldöntési időmérték aszerint sorolja be a feladatokat, hogy mekkora időegységeken értelmezhető egy becslés. Ez egy relatív mérték. Például a dallamfelismerés eldöntési időmértékére azt mondhatjuk, hogy alacsony, mert egy

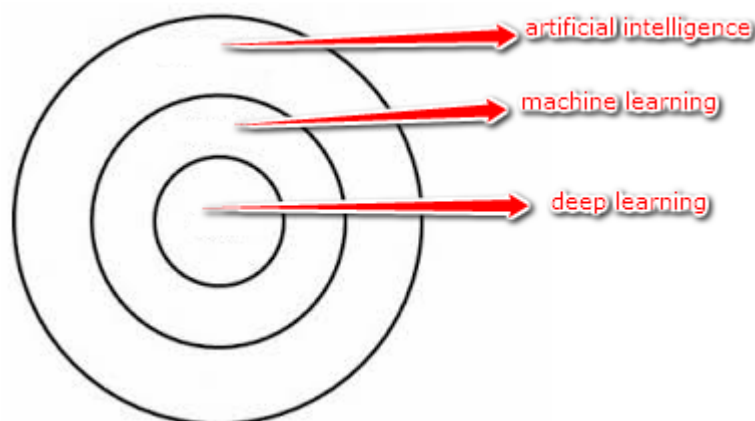
felismert dallam jó eséllyel nem fedi le az egész zenét. Másik kifejezéssel azt mondhatjuk, hogy ez egy időben változó, azaz dinamikus tulajdonság. Ellenben a tempó általában állandó értékű az egész zenében, így a teljes zeneszámot fel tudjuk címkézni egy adott tempóval. Erre azt mondjuk, hogy eldöntési időmértéke relatív magas, azaz ez egy statikus tulajdonsága a zenének.[15]

A hangszerfelismerést tekinthetjük statikus, illetve dinamikus feladatnak is a probléma megközelítésének függvényében. Dinamikus, ha erős címkézést szeretnénk megvalósítani, tehát arra vagyunk kíváncsiak, hogy adott időpillanatban éppen milyen hangszerek szólalnak meg. Gyenge címkézés esetén viszont a feladat statikussá válik. Ebben az esetben az egész zenére vetítve szeretnénk címkéket kapni egyes hangszerek jelenlétével, vagy a többi hangszerrel szembeni dominanciájával kapcsolatban.

TODO több MIR feladat

2.2. Mesterséges intelligencia

A mesterséges intelligencia egy általános fogalom az emberi gondolkodás számítógéppel való reprodukálására történő módszerekre. Ahogy arról a bevezetésben is szót ejtettem, a MIR tudományág gyakran használ mesterséges intelligencián alapuló megoldásokat. A továbbiakban két mesterséges intelligenciát megvalósító módszert mutatok be röviden: elsőként a gépi tanulást, majd a mély tanulást, amely a gépi tanulás egy ágazata és napjaink meghatározó trendje. [19]



2.2. ábra. Egyes fogalmak közti tartalmazás szemléltetve, forrás: [19]

2.2.1. Gépi tanulás (machine learning)

A gépi tanulás tehát a mesterséges intelligencia megvalósításának egy módszere. Lényege, hogy explicit utasításokat tartalmazó program helyett a bemeneti adatokat egy tanító algoritmusnak adjuk át. Ha elég mennyiségű és minőségű adatot szolgáltatunk a tanításhoz, akkor a modellünk

Hagyományos gépi tanulásról beszélünk, amikor

//TODO

2.2.2. Mély tanulás (deep learning)

A mély tanulás gyakorlatilag a gépi tanulás egy részhalmaza.

//TODO miben más mint az ML, architektúrák stb

3. fejezet

Adathalmaz

Ebben a fejezetben a munkám kapcsán felkutatott és alkalmazott adathalmazokról lesz szó. Egy deep learning megoldás tervezésének első lépéseként érdemes egy alkalmas kiinduló adathalmazt kiválasztani. Ezt aztán a modell tanítására és tesztelésre használjuk.

3.1. Kiválasztási szempontok

//TODO ismir dataset gyűjtemény pl. többszólam, ingyenesen elérhető, szerteágazó (not biased), stb gyenge címkézés

3.2. Philharmonia Orchestra

Kutatásom első fázisában a Philharmonia Zenekar ingyenesen elérhető hangmin-ta könyvtárát használtam fel. Ebben egyszólamú mintákat találunk. A minták a főkönyvtáron belül a bennük megszólaló hangszer nevével megegyező könyvtárban találhatóak, ez biztosítja a címkéket.

//TODO tulajdonságai

3.3. OpenMIC

A többszólamúság bevezetését a kutatásomban az OpenMIC [20] adathalmaz felhasználásával értem el. //TODO openmic cikk alapján

4. fejezet

Módszertan

4.1. Előfeldolgozás

//TODO adattisztítás, reprezentáció változtatás

4.2. Architektúra

//TODO felhasznált deep learning architektúra(?)

4.3. Megvalósítás

//TODO implementáció részletei

4.3.1. Könyvtárak

//TODO fejlesztői könyvtárak, függőségek - nem biztos hogy kell

5. fejezet

Kísérletek, eredmények

Lorem ipsum

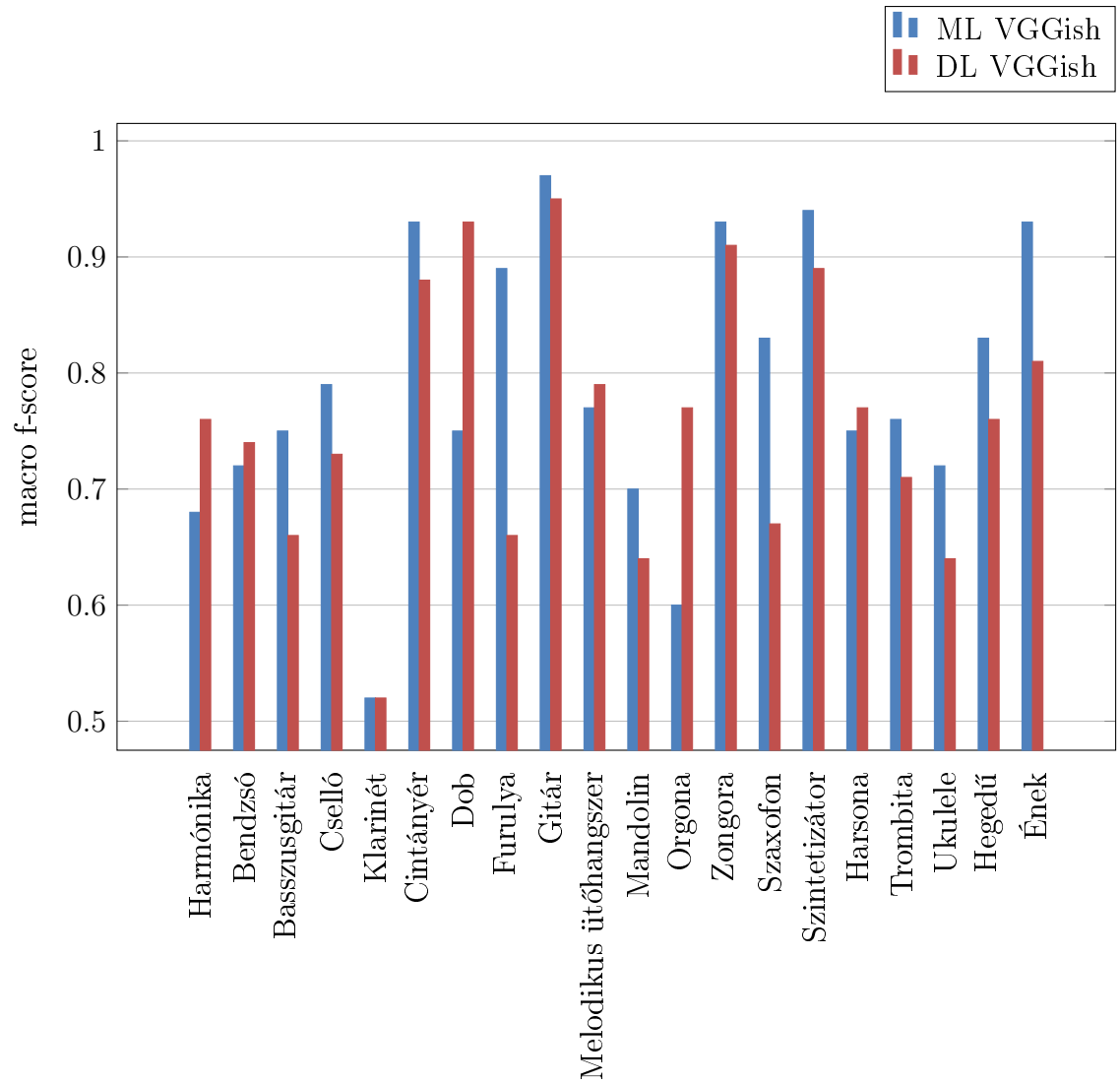
5.1. Mérőszámok

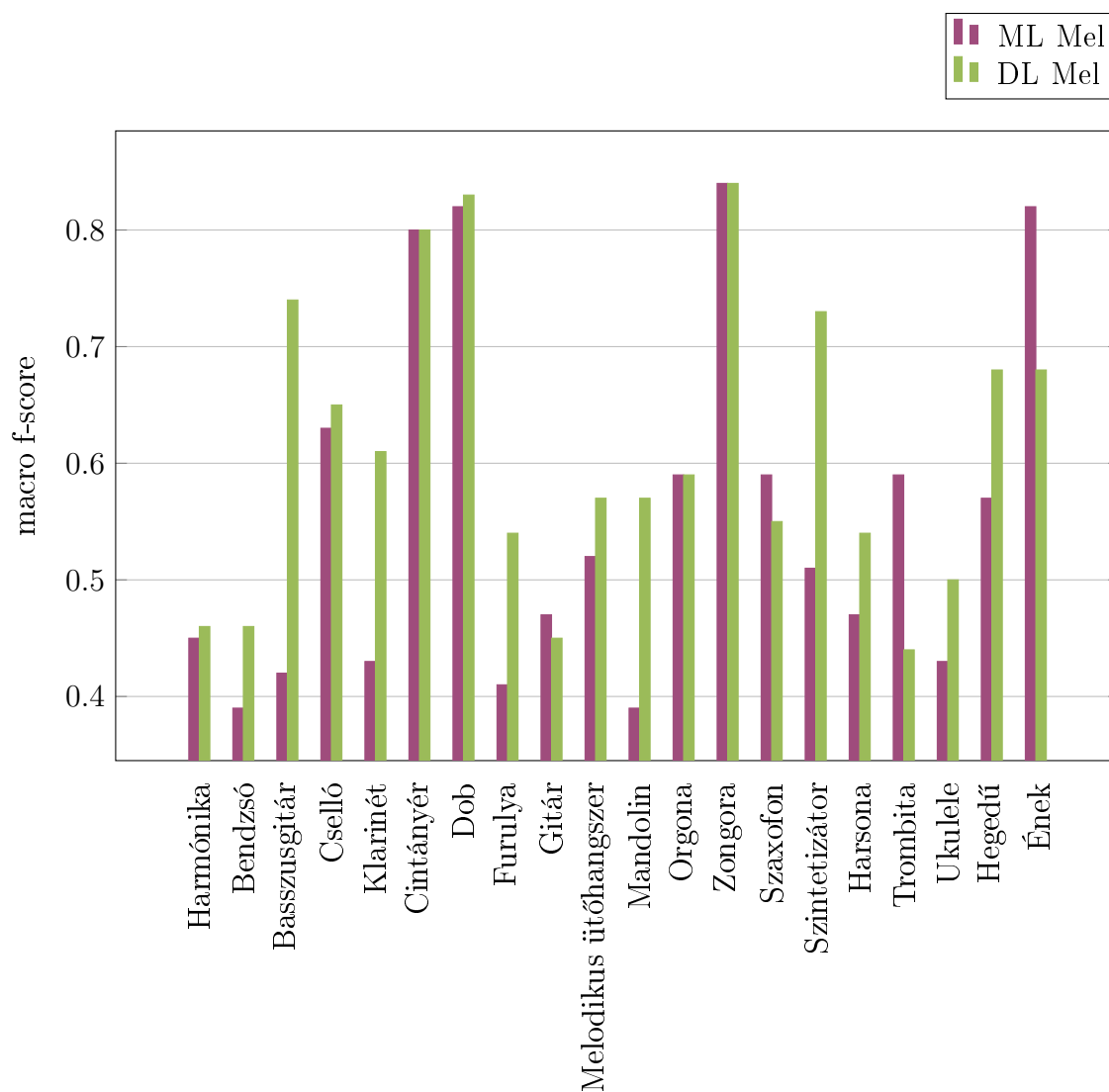
hogyan tudjuk megmérni a modell teljesítményét, pontosság, tanítás ideje, stb...

A következő metrikákat vizsgáltuk:

- Pontosság (accuracy) - a modell az adott lépésben a bemeneti adatok hány százalékára adott helyes kimenetet?
- Veszteség (loss) - a veszteségfüggvény eredménye. A modell predikcióinak a valóságtól való eltérését összeadva kapjuk meg.
- Precizitás (precision) -
- Felidézés (recall) -
- F1 érték (f1 score) - Ennek a súlyozott átlaga a mérvadó.

5.2. Eredmények





5.2.1. próba1

5.2.2. próba2

5.2.3. próba3

5.2.4. próba4

5.3. Összehasonlítás

5.3.1. saját próbálkozásaim

5.3.2. más munkákkal (hagyományos ML)

6. fejezet

Összegzés, kitekintés

Irodalomjegyzék

- [1] Florian Eyben és tsai. “Universal Onset Detection with Bidirectional Long Short-Term Memory Neural Networks.” *Proceedings of the 11th International Society for Music Information Retrieval Conference* (Utrecht, Netherlands). Utrecht, Netherlands: ISMIR, 2010. aug., 589–594. old. DOI: 10.5281/zenodo.1417131. URL: <https://doi.org/10.5281/zenodo.1417131>.
- [2] Philippe Hamel és Douglas Eck. “Learning features from music audio with deep belief networks”. *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*. Utrecht, The Netherlands, 2010. aug., 339–344. old.
- [3] Keunwoo Choi és tsai. “A tutorial on deep learning for music information retrieval”. *arXiv preprint arXiv:1709.04396* (2017).
- [4] Keith Dana Martin és Youngmoo E. Kim. “Musical instrument identification: A pattern-recognition approach”. 1998.
- [5] Judith Brown. “Computer Identification of Musical Instruments Using Pattern Recognition With Cepstral Coefficients as Features”. *The Journal of the Acoustical Society of America* 105 (1999. ápr.), 1933–41. old. DOI: 10.1121/1.426728.
- [6] A. Eronen és A. Klapuri. “Musical instrument recognition using cepstral coefficients and temporal features”. *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*. 2. köt. 2000, II753–II756 vol.2.
- [7] Jeremiah Deng, Christian Simmermacher és Stephen Crane field. “A Study on Feature Analysis for Musical Instrument Classification”. *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the*

- IEEE Systems, Man, and Cybernetics Society* 38 (2008. máj.), 429–38. old.
DOI: 10.1109/TSMCB.2007.913394.
- [8] Daulappa Bhalke, C. Rao és D. Bormane. “Automatic musical instrument classification using fractional fourier transform based- MFCC features and counter propagation neural network”. *Journal of Intelligent Information Systems* 46 (2015. máj.). DOI: 10.1007/s10844-015-0360-9.
- [9] Juan José Burred, Axel Roebel és Thomas Sikora. “Dynamic Spectral Envelope Modeling for Timbre Analysis of Musical Instrument Sounds”. *Audio, Speech, and Language Processing, IEEE Transactions on* 18 (2010. ápr.), 663–674. old.
DOI: 10.1109/TASL.2009.2036300.
- [10] J. Eggink és Guy Brown. “A missing feature approach to instrument identification in polyphonic music”. 5. köt. 2003. nov., 49–. old. ISBN: 0-7803-7850-4.
DOI: 10.1109/ICASSP.2003.1200029.
- [11] Dimitrios Giannoulis és Anssi Klapuri. “Musical Instrument Recognition in Polyphonic Audio Using Missing Feature Approach”. *Audio, Speech, and Language Processing, IEEE Transactions on* 21 (2013. szept.), 1805–1817. old.
DOI: 10.1109/TASL.2013.2248720.
- [12] Jayme Barbedo és George Tzanetakis. “Musical Instrument Classification Using Individual Partial”. *Audio, Speech, and Language Processing, IEEE Transactions on* 19 (2011. febr.), 111–122. old. DOI: 10.1109/TASL.2010.2045186.
- [13] Wenxin Jiang és Zbigniew Ras. “Multi-label automatic indexing of music by cascade classifiers”. *Web Intelligence and Agent Systems* 11 (2013. ápr.), 149–170. old. DOI: 10.3233/WIA-130268.
- [14] Eric J. Humphrey, Juan Pablo Bello és Yann LeCun. “Moving Beyond Feature Design: Deep Architectures and Automatic Feature Learning in Music Informatics.” *Proceedings of the 13th International Society for Music Information Retrieval Conference* (Porto, Portugal). Porto, Portugal: ISMIR, 2012. okt., 403–408. old. DOI: 10.5281/zenodo.1415726. URL: <https://doi.org/10.5281/zenodo.1415726>.

- [15] Keunwoo Choi és tsai. “A Tutorial on Deep Learning for Music Information Retrieval”. *CoRR* abs/1709.04396 (2017). arXiv: 1709.04396. URL: <http://arxiv.org/abs/1709.04396>.
- [16] Peter Li, Jiyuan Qian és Tian Wang. “Automatic instrument recognition in polyphonic music using convolutional neural networks”. *arXiv preprint arXiv:1511.05520* (2015).
- [17] Yoonchang Han, Jaehun Kim és Kyogu Lee. “Deep convolutional neural networks for predominant instrument recognition in polyphonic music”. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 25.1 (2016), 208–221. old.
- [18] J. Stephen Downie. “Music information retrieval”. *Annual Review of Information Science and Technology* 37.1 (2003), 295–340. old. DOI: 10.1002/aris.1440370108. eprint: <https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/aris.1440370108>. URL: <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/aris.1440370108>.
- [19] Dr. Michael J. Garbade. *Clearing the Confusion: AI vs Machine Learning vs Deep Learning Differences*. URL: <https://towardsdatascience.com/clearing-the-confusion-ai-vs-machine-learning-vs-deep-learning-differences-fce69b21d5eb>. Felkeresve: 2020. 05. 04.
- [20] Eric Humphrey, Simon Durand és Brian McFee. “OpenMIC-2018: An Open Data-set for Multiple Instrument Recognition.” *ISMIR*. 2018, 438–444. old.