Alkalmazott Matematikai Lapok **37** (2020), 87–101.
DOI: 10.37070/AML.2020.37.1.04

KÉZZEL ÍRT SZÁMJEGYEKET FELISMERŐ NEURONHÁLÓ ROBUSZTUSSÁGI VIZSGÁLATA

BISCHOF BARBARA HAJNALKA, KISS ATTILA ELEMÉR

A cikkben egy speciális adatbányászati algoritmust, nevezetesen a kézzel írt számjegyeket felismerő neurális hálót vizsgáljuk, miközben az adathalmazt egyre zajosabbá tesszük véletlen torzítások hozzáadásával. A tanító és a teszt adatok zajossá tételéhez többféle módszert is alkalmazunk. Részletesen elemezzük, hogyan hat a zaj az osztályozó algoritmusra. Összefüggést találunk a felismerés pontossága és aközött, hogy a teszt és a tanító adathalmazok milyen mértékben és milyen módon tartalmaznak zajt.

1. Bevezetés

Az adatbányászat olyan technológia, amely képes arra, hogy elemezze a nyers adatokat információ szerzés céljából. Az elnevezés megtévesztő, hiszen nem adatot, hanem számunkra hasznos információt, új és eddig rejtett összefüggéseket keresünk egy nagy adathalmazban.

Manapság adatok millióit tároljuk különböző adatbázisokban, melyeknek egy igen jelentős részét soha nem hasznosítjuk. Emiatt jelentősen megnőtt az igény mind a piaci élet résztvevői, mind a kutatók felől, a hatalmas adatbázisokból való információ keresésére. Ennek két fő oka van: egyrészt a növekvő versenyhelyzet miatt az üzleti szféra szereplőinek szüksége van az adatbázisokban megbújó hasznos információkra, így ez a fokozódó igény növekvő kutatói beruházásokat indukált. Másrészt az adatbányászat a maga multidiszciplináris (több tudományágat érintő) voltával attraktív terület számos kutató számára.

A sikeres adatbányászat alapfeltételei közt említhetjük értelemszerűen a nagy mennyiségű adatot, hiszen minél nagyobb az adatmennyiség, annál biztosabban tudjuk kizárni bizonyos összefüggések esetiségét, azaz annál kisebb az esélye, hogy a talált összefüggés csupán a véletlen eredménye.

További alapfeltétel az adatok tisztasága. A zajok, illetve hibás bejegyzések jobb esetben csak nehezítik az adatbányászatot, rosszabb esetben azonban hamis eredményekhez vezetnek. Tekintsünk most el azoktól az esetektől amikor az adatokat szándékosan torzítjuk, például személyes adatok védelmének érdekében.

A cikkben egy egyszerűbb kézírásfelismerő program segítségével mutatjuk be, hogy az adatok különböző módon való torzítása esetén a neuronháló mennyire ismeri fel az adott karaktert. Összefüggéseket mutatunk a felismerés pontosságára az alapján, hogy a teszt és a tanító adathalmaz milyen mértékben és módon tartalmaz zajt.

2. Kapcsolódó munkák

A tíz legnépszerűbb adatelemzéssel, klaszterezéssel és statisztikával foglalkozó algoritmus leírását a [8] publikációban találhatjuk meg. További algoritmusok részletesebb leírásával és egymástól eltérő adatbázisra való tesztelésével, illetve ezen eredmények összehasonlításával és elemzésével [3] foglalkozik.

A kézírásfelismerő programokról általánosságban, illetve az ehhez kapcsolódó kérdésekről [6]-ban olvashatunk részletesebben. [5] egy olyan új algoritmust mutat be, mely kézzel írott számok offline felismerésére alkalmas egy egyszerű többrétegű neurális hálózat felhasználásával, a hálózat a hasonló számok hatékony osztályozására alkalmas. Az összetett mintázatfelismerési problémák megoldására a [2] cikkben három összetett neuronhálózati osztályozót mutatnak be. A beszéd-, illetve kézírásfelismerésben alkalmazott mély neurális hálókról [7]-ben olvashatunk, a cikk bemutat egy olyan módszert, melynek segítségével elérhetjük, hogy némi zaj hozzáadásával a program rosszul osztályozzon adatokat.

Részletesebb és átfogóbb magyar nyelvű szakirodalom [4], az adatbányászat alapvető fogalmaival és főbb területeivel foglalkozik.

3. Elméleti háttér

A mesterséges neuronhálózat egy biológiai ihletésű program, ami a biológiai neurálisháló néhány tulajdonságát modellezi. Ezen modelleket természetesen nemcsak a biológiában, hanem számos más területen alkalmazzák főként tanító rendszerként. Leggyakoribb példája a képfelismerés, vagyis kézírásos vagy digitális szöveg szkennelésétől egészen az arcfelismerésig.

A tanulási technika szempontjából megkülönböztetünk ellenőrzött, illetve nemellenőrzött típusú tanulást. A kutatásunk során is alkalmazott ellenőrzött (felügyelt tanulás) esetében a rendszer számára nagy számú tanító mintapont párok (be- és kimeneti értékek) állnak rendelkezésre és a tanítás ezeken az ismert összerendelt mintapárokon alapul. Míg nemellenőrzött tanításnál címkézetlen tanítópontjaink vannak, így a hálózatnak kívánt válaszok ismerete nélkül kell valamilyen viselkedést kialakítania, a környezetből azonban nincs semmiféle visszajelzés, ami a hálózat viselkedésének helyességére utalna.

Karakterfelismerő rendszereknél megkülönböztetünk online és offline felismerést, ez a tulajdonsága arra utal, hogy a feldolgozás azonnal, közvetlenül a betűk, vagy szó beírása után, vagy passzív módon, jelentősen később történik. Az online írásfelismerők jellemzően egy úgynevezett digitális tinta (digital ink) technikát alkalmaznak, ahol a beviteli eszköz mozgásának folyamata kerül feldolgozásra, vagyis rendelkezésre áll az írás képzésének módja is. Ezzel ellentétben a cikkben is tárgyalt offline technikát használó karakterfelismerők, az írás befejeztével kapott betűk képét használják fel, azon felül nem rendelkeznek további információval.

A karakterfelismerő algoritmusnak két alapvető eleme van, az úgynevezett tulajdonságkinyerő (feature extractor) és az osztályozó (classifier). A tulajdonság analízis meghatározza azon jellegzetességeket, amikkel a karakter rendelkezik, majd ezt küldi el az osztályozónak. Az egyik leggyakoribb osztályozó eljárás a mintafelismerés, ebben az esetben az egyes pixelek az adott karakterkép sajátosságai. Az osztályozás során az algoritmus képkockánként összehasonlítja a bemeneti képet a különböző karakter-osztályok mintáival. Ennek eredményeként egy mérőszámot kapunk, amely megadja, hogy mennyire hasonló a bemenet és az adott minta, az eredménye az a karakter lesz, amihez a minta a legjobban hasonlított.

4. Kísérletek

A vizsgálatainkhoz olyan neurális hálózatot építettünk ki, amely képes a kézírásos számok helyes azonosítására. Ehhez a széles körben elterjedt MNIST adatbázist [1] használtuk, mely kézírásos számjegyeket tartalmaz.

Az MNIST adatbázisban a tanító adathalmaz (train) 60 ezer mintát, míg a tesztkészlet 10 ezer képet foglal magába. A képek halmaza tulajdonképpen egy nagyobb adatbázis (NIST) része, amely közel 250 író példáit tartalmazza (a képeket 8 biten ábrázoljuk, vagyis legfeljebb 256-féle árnyalatot látunk). Ezen képeket méretnormalizálták, továbbá a számjegyeket a rögzített kép közepére helyezték. Az általunk használt képadatok 28×28 szürkeárnyalatos képpont (összesen 784 képpont) formájában vannak rögzítve, címkével együtt a kép helyes azonosításához.

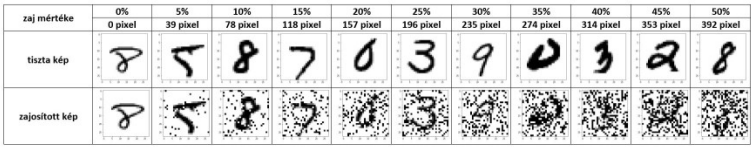
A karakterfelismerő programhoz importálnunk kell a Kereas-t, mely egy a Python programozási nyelvhez elérhető könyvtár, ami Tensorflow-ra, Theano-ra vagy CNTK-ra épül és kifejezetten mély tanuláshoz és neurális hálózatok gyors definíciójához használható. Használata előnyös, mert így nem kell manuálisan kódolni a lineáris algebrát, valamint a szükséges aktivációs függvényeket (activation function) és optimalizálókat.

Egy neuronhálózat elrendezése véletlenszerű, mivel annyi rejtett réteggel rendelkezik, amennyire szükség van, és az egyes rétegeken is eltérő számú neuronok lehetnek. Az általunk felépített neuronhálózatba 784 képpont fog bemenni, ezen pixeleket egy 512 neuronból álló rejtett rétegnek adjuk át, amely ezt 10 neuronnak adja kimenetként (minden számjegyre egyet).

Kutatásunkhoz szükségünk volt különböző módon és mértékben zajosított képek előállítására, ehhez alapvetően 5 különböző módszert használtunk, melyek közös jellemzője, hogy a zaj mértéke paraméterezhető volt, így segítségükkel több adathalmazt is elő tudtunk állítani. A következőkben ezen zajgeneráló technikákat mutatjuk be részletesebben.

4.1. Első módszer - véletlenszerű zaj

Első esetben minden kép esetén adott számú képpontot választottunk ki (ezt a számot az összes képpont számának és az adott adathalmaz zaj százalékának a szorzata adja), véletlenszerűen, egyenletes eloszlással, majd a képpontot értékétől függően, 0-ra vagy 255-re állítottuk át azt. Ha a képpont értéke nagyobb volt, mint 128, akkor 0-ra, ellenkező esetben pedig 255-re, így biztosítva, hogy a képpont mindig változzon (legfeljebb 255-tel és legalább 128-cal nő vagy csökken az értéke). A módosított képpontok száma minden kép esetén megegyezik (elhelyezkedésük azonban eltérő) egy adathalmazon belül. Összesen 50 adathalmazt állítottunk elő ezzel a módszerrel, a zaj mértéke pedig 1-től 50 százalékig terjed, ami minimum 8, maximum 392 zajos képpontnak felel meg.

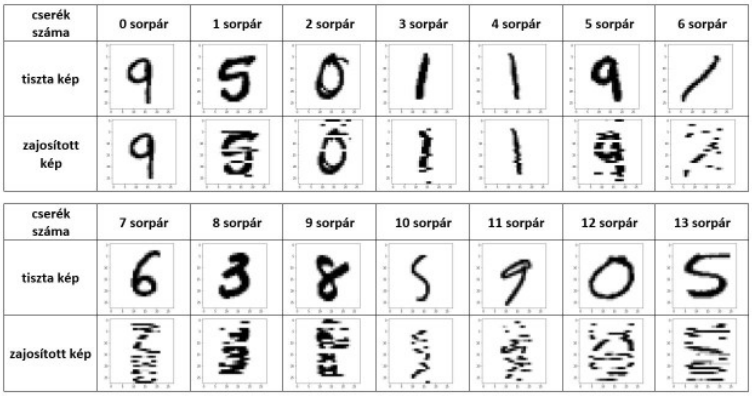


1. ábra. Zajosított kép - véletlenszerű zaj hozzáadásával

Az 1. ábra mutatja, hogy egy-egy kép mennyire tér el az eredetitől, ha adott százaléknyi zajt adunk hozzá. A karakter 25 és 30 százalékos zaj esetén is könnyedén felismerhető, de nagy valószínűséggel még 50 százaléknyi torzításnál is meg tudjuk mondani, hogy milyen szám szerepel a képen.

4.2. Második módszer - sorok cseréje

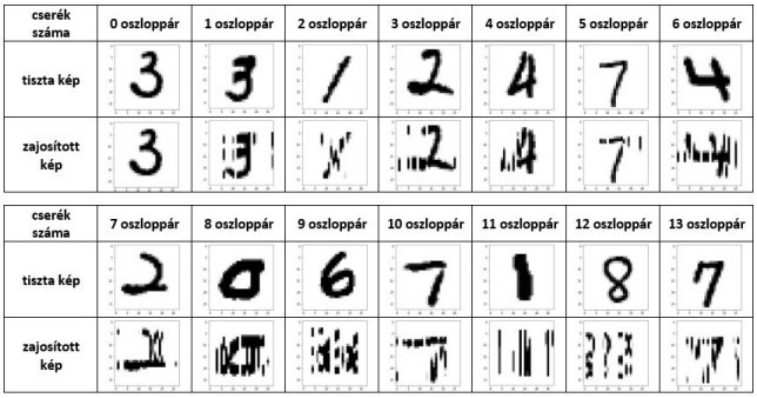
A második módszernél szintén véletlenszerűen egyenletes eloszlással választottunk két sorindexet (1 és 28 között), majd a kiválasztott indexek alapján megcseréltük a kép két sorát. Minden képnél más sorindexet jelöltünk ki, de egy adatkészleten belül a megcserélt sorpárok száma mindig azonos. Továbbá fontos megjegyezni, hogy a cserék egymás után hajtódtak végre (így kis valószínűséggel, de lehetséges, hogy valójában nem változott a kép). A 2. ábrán látható, hogy hogy néz ki az eredeti, illetve adott számú sorpár értékeinek felcserélése után a kép.



2. ábra. Zajosított kép - sorok felcserélésével

4.3. Harmadik módszer - oszlopok cseréje

Harmadik esetben a másodikhoz hasonló technikát választottunk, azzal a különbséggel, hogy nem a sorokat, hanem az oszlopokat cseréltük fel a képeken, ennek eredményét a 3. ábrán láthatjuk.

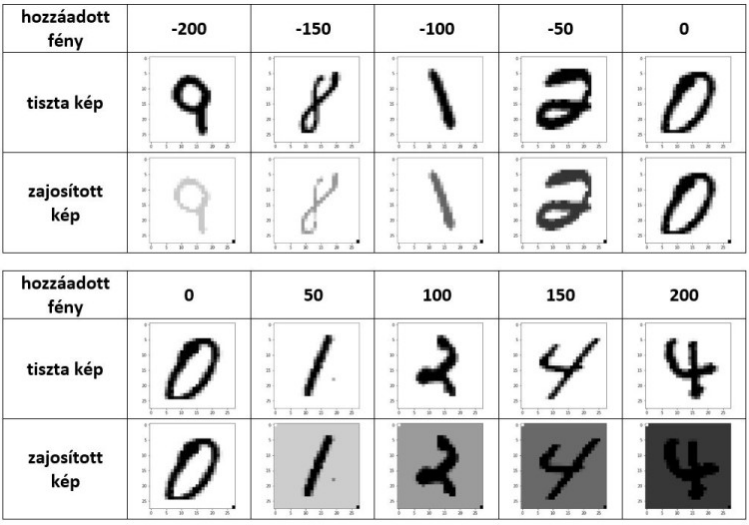


3. ábra. Zajosított kép - oszlopok felcserélésével

Amint azt a 2. és a 3. ábra jól mutatja, a számok aránylag jól felismerhetők vagy kikövetkeztethetők. Ebben jelentős szerepet játszik, hogy a képen belül a szám középre van igazítva, ezáltal a kép szélén elhelyezkedő sorok és oszlopok nem befolyásolják nagy mértékben a képek olvashatóságát, így ezeket egymással megcserélve a karakter továbbra is könnyedén felismerhető.

4.4. Negyedik módszer - fény változtatása

A negyedik módszer esetében egy adathalmazon belül minden képet egységesen világosítottuk vagy sötétítettük. Minden pixelhez hozzáadtunk, egy előre megadott értéket (-200, -150, -100, -50, 0, 50, 100, 150, 200) és ha az így kapott érték a [0, 255] intervallumon kívülre esett, akkor azt az intervallum megfelelő végpontjával helyettesítettük. Az így generált képeket a 4. ábra szemlélteti (minden kép bal felső pixelét fehérre (0), míg a jobb alsó pixelt feketére állítottuk (255), jobban szemléltetve a fény változását).

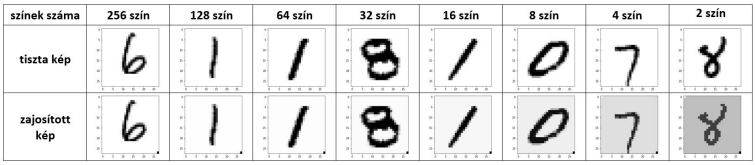


4. ábra. Zajosított kép - fény értékének változtatásával

Általánosságban elmondható, hogy a -100 és 100 között szinte biztosan és egyértelműen felismerhető a szám, míg -200, -150, 150 és 200-as érték hozzáadása esetén nehezebben tudjuk csak beazonosítani.

4.5. Ötödik módszer - színek számának változtatása

Az ötödik technika a kép intenzitásának változtatásán alapul. Az eredeti képeket 8 biten ábrázoltuk, így 256 különböző árnyalatot tudtunk megkülönböztetni. A következőkben ezt módosítjuk oly módon, hogy minden pixelt annak értékétől függően hozzárendelünk egy csoporthoz és az egy csoporthoz tartozó pixeleket ugyanarra a színre állítjuk be (az intervallum középső elemének színére), ezáltal 7 új tanuló és teszt adatkészletet hozunk létre, ahol az adathalmazokban a színek száma: 128, 64, 32, 16, 8, 4 és 2.



5. ábra. Zajosított kép - intenzitás változtatással

Az 5. ábra szemlélteti az eredeti, illetve a torzított képet. Ahogy látjuk, 16 szín esetén nem feltétlen tudjuk az eredeti és a zajosított képet megkülönböztetni egymástól, ezenfelül elmondhatjuk, hogy még 2 szín esetén sem romlott jelentősen az olvashatóság, és a karakter szépen kivehető.

5. Eredmények

Az előzőleg bemutatott zajgeneráló technikák segítségével különböző tanuló és teszt adathalmazokat készítettünk, majd egy-egy tanuló adathalmaz segítségével létrehozott neurális hálót minden (az adott módszer segítségével előállított) teszt adathalmazzal teszteltünk.

5.1. Első módszer - véletlenszerű zaj

train adatokban a zaj		teszt adatokban a zaj										
	db	0	39	78	118	157	196	235	274	314	353	392
db	%	0	5	10	15	20	25	30	35	40	45	50
0	0	98,2	88,8	65,6	48,0	37,7	30,0	25,5	21,6	19,3	17,1	16,1
8	1	98,0	89,6	67,1	51,0	39,6	32,4	27,0	23,0	19,8	17,5	16,2
16	2	97,6	93,7	71,8	47,5	31,5	22,7	17,5	14,9	12,8	11,6	11,0
24	3	98,0	95,6	83,6	61,7	43,1	31,8	24,9	21,7	18,5	17,2	15,9
31	4	97,6	96,4	89,1	73,1	53,9	39,8	29,5	23,8	20,3	18,1	16,7
39	5	97,2	96,5	91,5	79,6	62,0	48,0	36,4	29,8	24,2	20,2	18,4
78	10	96,5	96,5	94,5	91,1	84,3	74,1	61,7	50,7	41,2	33,3	28,0
118	15	96,8	96,5	95,1	93,3	90,0	84,9	78,1	70,2	60,9	52,6	46,3
157	20	95,5	95,6	94,7	93,2	91,3	87,6	83,1	76,9	69,9	61,6	55,1
196	25	95,7	95,2	94,7	93,1	91,3	88,5	85,1	80,0	73,8	66,6	60,8
235	30	94,4	93,9	93,0	92,1	90,2	87,7	84,2	80,1	74,5	68,3	63,1
274	35	94,5	93,6	92,8	91,9	90,2	87,6	84,6	80,9	77,0	70,7	66,1
314	40	93,7	93,0	92,4	91,3	89,6	87,7	85,3	80,8	77,0	71,2	66,3
353	45	91,3	90,7	90,3	89,1	87,9	85,8	83,5	80,2	76,5	70,6	66,4
392	50	91,9	91,2	90,1	89,2	87,3	86,0	83,1	79,1	75,5	70,4	65,5

1. táblázat. Felismerés pontossága a véletlenszerűen hozzáadott zaj hatására

Az 1. táblázatban találjuk az első zajgeneráló módszerrel készült képek esetén a felismerés pontosságának eredményeit (jobb olvashatóság érdekében a táblázat nem tartalmaz minden eredményt). A sorok megadják, hogy a tanító adathalmazban egy kép esetén hány százalék a zaj (illetve hogy ez hány darab pixelt jelent), míg az oszlopok a teszt adatokra vonatkoznak. Egy adott sor egy adott neuronhálót jelent, melyet különböző mértékben zajosított adatokkal teszteltünk.

A legjobb eredményt (98,2 százalék) értelemszerűen abban az esetben értük el, amikor a tanító adathalmaz és a teszt adathalmaz sem tartalmazott zajt. Míg a legrosszabb értékeket (11,0 százalék) akkor kaptuk, amikor a modell által betanult képek minimális (2-3 százalék) zajt tartalmaztak és a teszt adatoknál pedig minden kép esetében a pixelek felének eltért a színe az eredetitől.

Azt mondhatjuk, hogyha a tanító adathalmaz képeinek zajossága X, a teszt adathalmaz képeinek zajossága Y, akkor igaz az alábbi összefüggés: ha $X \leq 25$ és $X - Y \geq 0$, vagy ha $X > 25$ és $X + Y \leq 50$, ebben az esetben igaz az, hogy a neuronháló legalább 90 százalékos valószínűséggel felismeri az adott karaktert.

5.2. Második és harmadik módszer - sorok és oszlopok cseréje

Az oszlopok, illetve sorok cseréjével előállított képek esetében, hasonló eredményeket kaptunk, ezeket a 2. és a 3. táblázatban látjuk. A legjobb érték a zaj nélküli adatokban keletkezett, ahogy azt az előző esetben is láttuk, míg a legrosszabb értéket akkor kaptuk, ha a teszt adathalmaz minden képében felcseréltünk 13 sort/oszlopot és a tanuló adatokat pedig nem zajosítottuk. Az így kapott legrosszabb értékek (40,6 százalék az oszlop és 32,6 százalék a sor cserék esetén) jelentősen jobbak, mint az első (százalékos zaj) módszerrel zajosított képeknél (11,1 százalék).

train adatokban a cserék száma \ teszt adatokban a cserék száma	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	98,3	89,7	83,8	77,3	70,2	63,6	58,5	52,3	47,9	43,8	40,7	37,2	34,5	32,6
1	97,9	96,4	95,0	93,0	91,3	87,6	84,9	80,8	77,0	72,9	70,2	65,8	63,1	60,5
2	97,6	96,5	95,5	94,0	92,2	89,6	87,0	83,6	79,7	76,2	73,2	69,0	66,3	63,5
3	97,7	96,8	95,7	94,7	93,2	91,1	88,8	85,7	82,6	78,6	76,2	71,9	69,6	66,5
4	97,3	96,4	95,6	94,8	93,7	91,9	90,0	87,6	84,6	81,8	78,9	75,2	72,4	69,8
5	96,9	96,4	95,6	94,7	93,8	92,4	90,9	88,5	86,5	83,7	81,1	77,7	75,6	73,0
6	96,8	95,9	95,2	94,6	93,6	92,4	90,6	88,6	86,0	83,9	81,6	78,4	75,1	73,4
7	96,7	95,8	95,5	94,3	93,9	92,6	91,5	89,7	87,7	85,6	83,8	80,5	78,5	76,0
8	96,3	95,5	94,8	94,4	93,9	92,4	91,4	89,9	87,5	85,8	84,4	81,5	80,2	77,3
9	96,0	95,5	94,8	94,2	93,5	92,2	91,7	90,1	88,0	86,4	85,4	82,8	81,1	79,1
10	95,7	95,0	94,3	93,8	93,4	92,3	91,0	89,7	88,0	86,9	85,6	83,2	81,5	80,0
11	95,3	94,5	93,9	93,4	92,8	92,2	90,8	89,8	88,2	86,9	85,9	83,9	82,0	80,1
12	95,3	94,6	94,1	93,2	92,5	91,7	91,1	89,9	88,6	87,5	86,4	83,6	82,6	81,1
13	94,9	93,9	93,3	92,8	92,1	91,4	90,2	89,4	87,3	86,8	85,2	83,9	82,0	81,0

2. táblázat. Felismerés pontossága adott számú sorpár felcserélésének hatására

Sorok esetén az alábbi összefüggés adja meg, hogy a felismerés pontossága hol nagyobb mint 90 százalék: ha a tanító adathalmaz képeiben a felcserélt sorok száma X, a teszt adathalmaz képeiben felcserélt sorok száma Y, akkor ha $X \leq 7$ és $X - Y \geq 0$, vagy ha $X > 7$ és $X + Y \leq 13$.

Oszlopok cseréje esetén az előzőhöz hasonlóan X a tanító adathalmazra, míg Y a tesztkészlet képeire vonatkozik, ekkor ha igaz, hogy $X - Y \geq 0$, akkor a neuronháló felismeri a karaktert 90 százalékos valószínűséggel.

train adatokban a cserék száma \ teszt adatokban a cserék száma	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	98,3	91,2	86,4	80,6	75,2	69,8	64,3	59,5	55,8	51,8	48,7	45,9	43,7	40,6
1	98,1	96,9	96,1	94,9	93,5	91,7	90,0	87,3	85,9	83,9	81,8	79,3	77,5	74,9
2	97,8	96,9	96,3	95,3	94,2	93,0	92,0	89,5	88,2	86,5	84,6	82,8	81,5	78,7
3	97,6	96,8	96,3	95,6	94,7	94,0	92,7	91,2	89,8	88,7	87,2	85,2	84,0	81,9
4	97,6	96,7	96,5	95,5	94,9	93,9	93,0	91,5	90,5	89,4	88,0	86,4	85,2	83,7
5	97,5	96,7	96,2	95,8	94,9	94,2	93,3	92,4	91,5	90,3	89,2	87,9	86,5	85,2
6	97,4	96,8	96,4	95,9	95,0	94,7	93,8	93,0	92,1	91,6	90,0	88,4	87,8	86,2
7	97,0	96,4	96,2	95,8	95,1	94,5	94,0	93,3	92,6	91,6	90,7	89,6	88,7	87,6
8	96,8	96,2	95,9	95,5	95,2	94,5	94,2	93,3	92,6	91,5	91,0	89,6	89,3	88,0
9	96,4	95,9	95,3	95,0	94,6	94,3	93,7	92,5	91,9	91,6	91,1	89,7	88,7	88,0
10	96,6	96,2	95,6	95,6	95,2	94,9	94,0	93,4	92,8	92,2	91,9	90,4	90,0	88,7
11	96,1	95,9	95,5	95,1	94,8	94,3	93,7	93,1	92,8	91,9	91,7	90,5	90,0	89,1
12	96,3	95,9	95,2	95,2	94,4	94,7	94,0	93,3	92,9	92,5	91,4	90,7	90,4	89,5
13	96,2	95,4	95,3	95,2	94,6	94,3	93,8	93,2	92,8	92,4	91,8	90,8	90,4	89,5

3. táblázat. Felismerés pontossága adott számú oszloppár felcserélésének hatására

Amint az a 2. és a 3. táblázatból is kiolvasható, adott számú oszlop felcserélése esetén számottevően jobb a karakter felismerésének pontossága (átlagosan 85,3 százalék), mint ugyanannyi sorpár felcserélése esetén (átlagosan 90,3 százalék). Ennek oka valószínűleg a karakterek elhelyezkedéséből adódik, hiszen a képek nagy részénél a tényleges betű egy kisebb téglalapban helyezkedik el a kép közepén. Így a tőle jobbra, illetve balra levő „szinte" fehér oszlopok felcserélése nem ront jelentősen az olvashatóságon.

5.3. Negyedik módszer - fény változtatása

A 4. táblázatban azt láthatjuk, milyen eredményeket kaptunk a felismerés pontosságára abban az esetben, amikor a fény erejét állítottuk az egyes képeken. Az előző kísérletekhez hasonlóan a legjobb eredményt akkor értük el, ha a tanuló adathalmaz és a teszt adathalmaz sem tartalmazott zajt, míg a legrosszabb értéket (10,28 százalék) abban a esetben kaptuk, ha a tanuló adathalmaz nem tartalmazott zajt, és a teszt adathalmaz képei pedig szemmel láthatóan sötétebbek (pixelek értékét 200-zal növeltük).

hozzáadott fény értéke a train adatokban \ hozzáadott fény értéke a teszt adatokban	-200	-150	-100	-50	0	50	100	150	200
-200	96,7	97,3	97,5	97,4	97,1	55,4	32,4	17,4	11,6
-150	96,5	97,6	97,9	97,9	97,7	51,9	31,0	18,8	11,2
-100	94,8	97,4	97,9	98,0	98,0	47,3	21,3	11,4	10,3
-50	92,8	97,1	97,8	98,0	98,1	47,0	20,5	10,5	10,3
0	91,6	96,9	97,6	98,0	98,1	46,3	16,9	10,5	10,3
50	95,0	96,2	96,6	96,9	97,1	97,5	85,2	39,8	13,6
100	93,9	95,0	95,5	95,7	95,8	96,8	96,7	89,4	36,7
150	90,1	90,6	90,8	90,7	90,3	92,0	93,0	92,1	64,7
200	80,6	80,1	79,6	79,1	78,2	79,7	81,3	83,8	82,5

4. táblázat. Felismerés pontossága a fény változtatásának hatására

Ha a tanító adathalmaz képeinek zajosságát X-szel és a teszt adathalmaz képeinek zajosságát Y-nal jelöljük, akkor igaz az alábbi összefüggés: ha $-200 \leq X \leq 150$ és $-200 \leq Y \leq 0$, vagy ha $50 \leq X, Y \leq 150$ és $Y - X \leq 0$, ilyenkor a karakter felismerésének pontosságának valószínűsége legalább 90 százalék.

5.4. Ötödik módszer - színek számának változtatása

train adatokban a színek száma \ teszt adatokban a színek száma	256	128	64	32	16	8	4	2
256	98,2	98,1	98,1	97,9	97,5	93,1	59,2	21,2
128	98,2	98,2	98,2	97,9	96,2	79,1	23,6	8,9
64	98,2	98,2	98,1	98,1	97,2	88,5	47,1	9,8
32	98,1	98,1	98,2	98,2	97,9	93,8	51,7	9,3
16	97,7	97,8	97,9	97,9	97,9	96,9	78,3	18,2
8	97,3	97,3	97,4	97,5	97,6	97,8	95,2	30,1
4	97,3	97,4	97,4	97,5	97,6	97,8	97,7	82,7
2	96,4	96,4	96,4	96,5	96,6	96,7	97,1	96,2

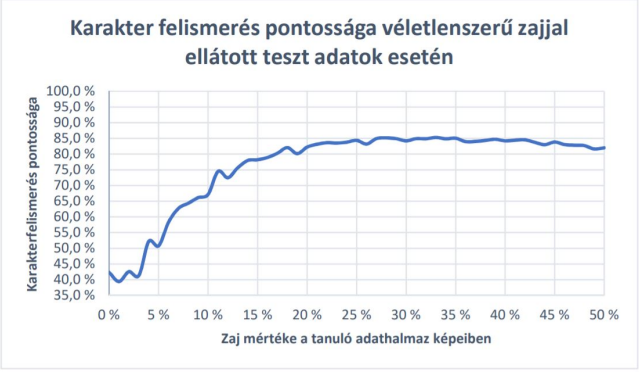
5. táblázat. Felismerés pontossága az intenzitás változtatásának hatására

A színek számának változtatására kapott eredményeket az 5. táblázat tartalmazza. Amint azt az 5. táblázaton is láthattuk, hogy nem romlott nagy mértékben az olvashatóság, így ennek megfelelően a karakter felismerésének pontosságára kapott eredmények is magasak (átlagukat tekintve 85,5 százalék - összehasonlításképp az első zajgeneráló módszerrel kapott eredmények átlaga 67,7 százalék).

Mivel a csupán 16 színt tartalmazó képet szabad szemmel szinte meg sem tudjuk különböztetni a 256 színt tartalmazó képtől, emiatt az 5. táblázatban látható legmagasabb értéket (98,2 százalék) több esetben is elértük. A felismerés pontosságára kapott legrosszabb eredményt (8,9 százalék) abban az esetben kaptuk, ha a tanuló adathalmaz képei 64 színből állt, míg a teszt készlet képei csupán 2-ből.

6. Következtetések

Az előző fejezetben tárgyalt kísérletek alapján azt mondhatjuk, ha egy adott mértékig zajos adatokat szeretnénk felismerni, akkor a legjobb módszer, ha a tanító adathalmazt is hasonló módon és mértékben zajosítjuk. Hiszen így tudjuk elérni a felismerés pontosságára a legjobb értéket. Azonban fontos megjegyezni, hogy ezekben az esetekben mind a teszt és mind a tanító adathalmaz elemei azonos mennyiségű zajt tartalmaztak, ami egy nagyon speciális, a valóságtól igencsak eltérő eset.

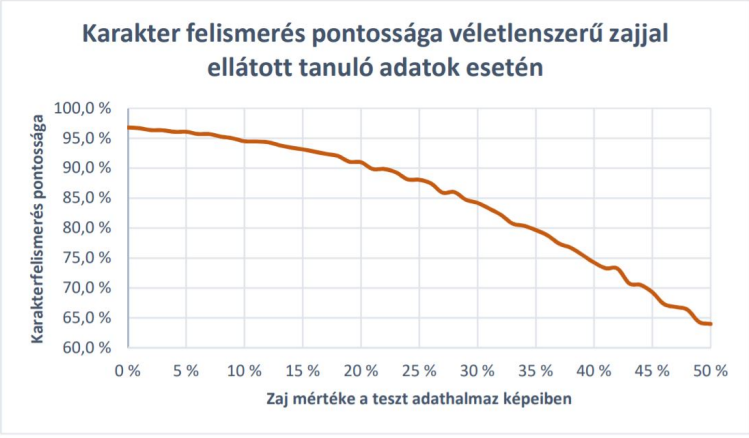


6. ábra. Felismerés pontossága véletlenszerűen zajosított tesztkészlet esetén

Mindezek miatt nézzünk egy olyan esetet, ahol több tanító adathalmazt hozunk létre, oly módon, hogy egy-egy halmazon belül azonos százaléknyi zajt tartalmazó képek szerepelnek (ezen tanító adathalmazok megegyezhetnek az 5. fejezetben tárgyalt tanító adathalmazokkal). Az így létrejött karakterfelismerő neuronhálók mindegyikét ugyanarra az egy teszt adatkészletre teszteljük, melyre az teljesül, hogy minden egyes kép esetén meghatároztunk egy véletlen számot (1 és 50 között) és ezzel a véletlenszámmal generáltunk zajt (az első - százalékos - zajgeneráló módszer segítségével) külön-külön mindegyik kép esetén, ahol a véletlenszám adta meg a zaj százalékos értékét. Ezáltal a tesztadathalmaz a valóságnak megfelelően eltérő mértékben tartalmaz zajt az egyes képeken.

Az így elért eredményeket a 6. ábrán láthatjuk, ahol a vízszintes tengely megadja, hogy a neuronháló milyen mértékű zajt tartalmazó képeken tanult, míg a függőleges tengely a karakter felismerésének valószínűségét adja meg. Amint azt láthatjuk, a legrosszabb esetben nagyjából 40 százalék valószínűséggel ismeri fel a képet a program, azonban egyetlen esetben sem éri el a 90 vagy annál nagyobb százalékot, átlagosan azt mondhatjuk, hogy a felismerés pontossága 75-77 százalék között mozog. Ezen értékek megegyeznek az 1. táblázatban a 25%-hoz tartozó oszloppal, vagyis ha a tesztkészlet minden képét azonosan rontottuk el 25 százalékban.

Ezenfelül tekintsünk egy további lehetőséget, amikor egy tanító adathalmazunk van, melynek minden képe más mennyiségű zajt tartalmaz és az ezen képek által felépített neuronhálót teszteljük különböző tesztkészletekkel, ebben az esetben a tesztkészleten belül a képek egyformán zajosak.



7. ábra. Felismerés pontossága véletlenszerűen zajosított tanító adathalmaz esetén

A 7. ábra diagramja mutatja, hogy ha egy neuronhálót építünk és azt különböző tesztkészletekre teszteljük (vízszintes tengely), akkor milyen pontossággal ismeri fel a program a karaktert. Az itt kapott értékek (legrosszabb esetben: 64,3 legjobb esetben: 96,8 átlagosan: 85,3 százalék) hozzávetőlegesen megegyeznek azzal az esettel, amikor a tanító adathalmaz minden képét azonosan zajosítottuk 25 százalékban, ez az 1. táblázatban a 25%-hoz tartozó sor.

Az azonos mértékben zajosított tesztkészlet, és a véletlenszerűen zajt tartalmazó között talált összefüggés miatt alátámasztást nyert az a megállapítás, hogy egy karakterfelismerő neuronháló felismerésének pontosságát javíthatjuk azáltal, hogy a tanító adathalmaz képeit zajosítjuk.

7. Kitekintés

A cikkben végzett kísérletek során a zaj mértékét és módját változtattuk, ehhez alapvetően 5 különböző módszert használtunk, ezenfelül hasznos lenne további technikák tesztelése, melyek életszerűbbek és jobban tükrözik a valóságot. Ilyen lehet például az elmosódott, homályos képek vagy a rossz fény beállításokkal készítünk fényképet (egyenletlen megvilágítás, erőteljes vaku) vagy elnyújtott/összezsugorított képek (nem megfelelő szögben tartott fényképezőgép esetén).

A továbbiakban még érdemes vizsgálnunk, hogy mivel tudjuk még jobban javítani a felismerés pontosságát, ez lehet esetleg a neuronháló alap beállításainak módosítása, vagyis a neuronok és idegrendszerek számának változtatása, vagy a tanuló adathalmaz méretének módosítása, továbbá a Keras által nyújtott különböző optimalizálók használata.

Ezek mellett azt is célszerű vizsgálni, hogy hogyan tudjuk eldönteni egy megadott képről, hogy milyen mértékben tartalmaz zajt, hiszen ennek az értéknek az ismeretében könnyen tudjuk úgy kalibrálni a neuronhálónkat, hogy minél nagyobb valószínűséggel ismerje fel az adott karaktert.

Köszönetnyilvánítás

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg (EFOP-3.6.3-VEKOP-16-2017-00002).

Továbbá köszönjük szépen az anonim bírálóknak a hasznos és értékes észrevételeiket, javaslataikat és megjegyzéseiket.

Hivatkozások

[1] The MNIST database of handwritten digits: *http://yann.lecun.com/exdb/mnist/*.

[2] Cho, Sung-Bae: *Neural-network classifiers for recognizing totally unconstrained handwritten numerals*, IEEE Transactions on Neural Networks, Vol. **8** No. **1**, pp. 43-53 (1997). DOI: 10.1109/72.55419

[3] Dogan, Neslihan, and Zuhal Tanrikulu: *A comparative analysis of classification algorithms in data mining for accuracy, speed and robustness*, Information Technology and Management, Vol. **14** No. **2**, pp. 105-124 (2013). DOI: 10.1007/s10799-012-0135-8

[4] Ferenc, Bodon: *Adatbányászati algoritmusok (2002).* http://www.cs.bme.hu/~bodon/magyar/adatbanyaszat/tanulmany/adatbanyaszat.pdf

[5] Lee, Seong-Whan: *Off-line recognition of totally unconstrained handwritten numerals using multilayer cluster neural network*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **18** No. **6**, pp. 648-652 (1996). DOI: 10.1109/34.506416

[6] Suen, Ching Y., et al.: *Handwriting recognition-the last frontiers*, Proceedings 15th International Conference on Pattern Recognition. ICPR-2000. IEEE, Vol. **4** (2000). DOI: 10.1109/ICPR.2000.902853

[7] Szegedy, Christian, et al.: *Intriguing properties of neural networks (2013)*, arXiv preprint 1312.6199 (2014).

[8] Wu, Xindong, et al.: *Top 10 algorithms in data mining*, Knowledge and information systems, Vol. **14** No. **1**, pp. 1-37 (2008). DOI: 10.1007/s10115-007-0114-2

Bischof Barbara Hajnalka 1998-ban született Körmenden. 2016-ban tett érettségit a körmendi Kölcsey Ferenc Gimnáziumban, majd az Eötvös Loránd Tudományegyetem Informatikai Karán folytatta tanulmányait Programtervező Informatikus szakon, ahol 2019-ben alapszakos diplomát szerzett. A mesterképzést 2019-ben kezdte Információs Rendszerek szakirányon szintén az ELTE-n. Emellett 2018-tól az SAP Hungary Kft.-nél dolgozik junior fejlesztő munkatársként.

Bischof Barbara Hajnalka

ELTE Eötvös Loránd Tudományegyetem Informatikai Kar
1117 Budapest, Pázmány Péter sétány 1/C
bisbarbi@caesar.elte.hu

Kiss Attila Elemér 1985-ben matematikusként végzett az Eötvös Loránd Tudományegyetemen. 1991-ben lett a matematikai tudomány kandidátusa. 2010-ben habilitált az informatikai tudományokból. 2010 óta az Eötvös Loránd Tudományegyetem Információs Rendszerek Tanszékének vezetője. Több mint 140 publikációja jelent meg, elsősorban adatbázisok, adatbányászat, mesterséges intelligencia, bioinformatika témakörökben. Doktori hallgatói közül eddig heten szerezték meg a doktori fokozatot. A kutatás mellett számos sikeres kutatás-fejlesztési, illetve ipari projektet vezetett.

Kiss Attila Elemér

ELTE Eötvös Loránd Tudományegyetem Informatikai Kar
1117 Budapest, Pázmány Péter sétány 1/C
kiss@inf.elte.hu

ROBUSTNESS TESTING OF NEURAL NETWORK FOR HANDWRITTEN DIGIT RECOGNITION

Barbara Hajnalka Bischof, Attila Elemér Kiss

The paper examines a special data mining algorithm, namely the neural network that recognizes digits, while making the data set increasingly noisy by adding random distortions. We analyze in detail how noise affects the classification algorithm. Using a simpler handwriting recognition program, we show how the neural network recognizes a given character when it distorts data in different ways (we used five different methods to noise the data). We find correlations for recognition accuracy based on the extent and way in which the test and train data sets contain noise.

Keywords: data mining, neural network, robustness, handwriting recognition.

Mathematics Subject Classification (2000): 68T05, 68T35.