VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
MATEMATINĖS INFORMATIKOS KATEDRA
BIOINFORMATIKOS SPECIALYBĖ

Rašysenos atpažinimas ir identifikavimas

Handwriting recognition and identification

Bakalauro baigiamasis darbas

Atliko: 4 kurso Bioinformatikos studentas
Edvinas Karenga

Darbo vadovas: Lekt. Irus Grinis

Recenzentas: Dr. Andrius Merkys

Vilnius 2021

TURINYS

Santrauka

Rašysena yra vienas iš žmogaus elgsenos biometrinių duomenų, o rašysenos atpažinimas yra viena iš kompiuterinės vaizdų analizės sričių. Šiame darbe nagrinėjama viena iš rašysenos atpažinimo užduočių – autorių atpažinimas, dažnai sutinkama kriminalistinėje dokumentų ekspertizėje. Darbe įgyvendinti du Siamo architektūros konvoliuciniai neuroniniai tinklai skirti statinio, nuo teksto nepriklausomo, autorių verifikavimo uždaviniui spręsti, t.y. atsakyti į klausimą ar tekstą parašė tas pats žmogus. Šių tinklų architektūros paremtos SigNet ir DeepWriter neuroniniais tinklais. Realizuoti modeliai apmokyti naudojant „IAM handwriting database“ duomenų rinkinį, o jų veikimui įvertinti ir palyginti atlikti du eksperimentai. Geriausi gautieji rezultatai, su vienu iš tinklų, siekė 95.05% verifikacijos tikslumą, su 3.34% vienodos klaidos tikimybe.

Raktiniai žodžiai: rašysenos atpažinimas, autorių atpažinimas, konvoliuciniai neuroniniai tinklai

Summary

Handwriting is one of human behavioral biometrics and handwriting recognition is a part of computer image analysis. This work covers one area of handwriting recognition - writer identification, which is commonly encountered in forensics document expertise. In this work, two Siamese architecture convolutional neural networks are implemented, for the static, text independent, writer verification task, to answer the question whether the text is written by the same person. The architecture of these networks is inspired by the SigNet and DeepWriter neural networks. Both models were trained on the ”IAM handwriting database”, and two experiments were performed to test and compare their performance. The best results, achieved with one of the networks, were 95.05% verification accuracy with 3.34% equal error rate.

Keywords: handwriting recognition, writer recognition, convolutional neural networks

Įvadas

Rašysena – žmogaus rašymo stilius, naudojant kokią nors rašymo priemonę [Pre]. Žmogaus rašysena yra vienas iš jo biometrinių duomenų, tačiau priešingai nei fiziniai biometriniai duomenys, kaip veido bruožai ar pirštų atspaudai, kurie yra įgimti, rašysena yra priskiriama prie elgsenos biometrijos, nes rašymo stilius yra tai, ką žmogus išmoksta.

Rašysenos atpažinimas yra viena iš vaizdų analizės sričių, leidžiančių kompiuteriui interpretuoti ranka rašytą tekstą tiek iš fizinių šaltinių, kaip popierinių dokumentų nuotraukos, tiek iš elektroninių šaltinių, tokių kaip patys paprasčiausi įrenginiai liečiamu ekranu ar specialiai tam pritaikytų skaitmeninių planšečių. Rašysenos atpažinimas apima platų spektrą su tuo susijusių uždavinių, pvz.: optinis simbolių atpažinimas, leidžiantis konvertuoti ranka rašytą tekstą į kompiuterinį tekstą, parašų verifikavimas, leidžiantis verifikuoti parašo autentiškumą, ar autorių atpažinimas, leidžiantis nustatyti teksto autorių.

Šiame darbe nagrinėjama viena iš rašysenos atpažinimo užduočių – autorių atpažinimas. Autorių atpažinimo sistemų tikslas yra identifikuoti ar verifikuoti autorių, pagal jo rašyseną. Viena iš šio uždavinio pritaikymo sričių yra biometrinės sistemos, identifikuojančios žmogaus tapatybę, kur rašysena gali būti būti vienintelis ar tik pagalbinis biometrinis duomuo, tačiau tai nėra taip plačiai išplitę, kaip veido, pirštų atspaudų ar akies rainelės atpažinimo sistemos, dėl gana didelės rašysenos variacijos, apsunkinančios tokių sistemų darbą, kai to paties žmogaus rašysena, skirtingų rašymų metu, gali skirtis gana stipriai. Vienas iš rašysenos atpažinimo sistemų privalumų yra tai, kad jų įgyvendinimas daugeliu atvejų yra pigesnis, nei pirštų atspaudų ar akies rainelės atpažinimo sistemų. Be biometrinių sistemų, rašysenos atpažinimas pritaikomas ir tokiose srityse kaip: automatizuota dokumentų ekspertizė, naudojama kriminalistikoje, kaip pagalbinė teksto autorių identifikavimo priemonė, senovės rankraščių klasifikacija [SCV09], padedanti skaitmenizuoti senovinius tekstus ir jų autorius, ar išmanieji susitikimų kambariai [LSB+06], leidžiantys ne tik konvertuoti susitikimo metu ranka rašytą tekstą į kompiuterinį dokumentą, bet ir atpažinti jo autorių.

Šio darbo tikslas yra įgyvendinti vieną iš rašysenos verifikavimo metodų, kuris atliktų dviejų tekstų palyginimą, ir leistų nustatyti ar tai yra to paties ar skirtingų žmonių rašytas tekstas. Šis metodas galėtų būti pritaikomas biometrinėse sistemose, verifikuojant žmogų, ar kriminalistikoje, užtikrinant tiriamojo teksto autentiškumą.

Darbo tikslui pasiekti išsikelti šie uždaviniai:

1. Susipažinti su rašysenos verifikavimo užduotimi, apžvelgti su ja susijusius darbus ir juose taikomus metodus.
2. Realizuoti vieną iš rašysenos verifikavimo metodų.
3. Su realizuotu metodu atlikti keletą eksperimentų ir palyginti gautuosius rezultatus.

1. Rašysenos atpažinimo apžvalga

Šiame skyriuje apžvelgiami rašysenos tipai, jos atpažinimo sistemų skirstymas pagal priklausomumą nuo teksto ir autoriaus. Taip pat apžvelgiami su rašysenos atpažinimu susiję darbai.

1.1. Rašysenos tipai

Visi skaitmenizuoti rašysenos duomenys yra skirstomi į dvi grupes: statinį tekstą (*angl. off-line*) ir dinaminį tekstą (*angl. online*) [IOO91]. Statinis tekstas yra tekstas parašytas ant popieriaus ranka, o vėliau skaitmenizuotas naudojant kameras ar skenavimo įrenginius. Dinaminis tekstas gaunamas rašymui naudojant elektroninius prietaisus, tokius kaip skaitmeninės planšetės, jutiminiai ekranai ar specialiai tam pritaikyti rašikliai.

Dinaminio teksto pranašumas prieš statinį, tai kad be teksto atvaizdo, yra suteikiama ir papildoma informacija, kaip rašymo greitis ir pagreitis skirtingose vietose, spaudimo stiprumas, rašiklio atkėlimo pozicijos ir kitos rašymo ypatybės.

Dėl daugiau ir įvairesnių dinaminio teksto suteikiamų duomenų, tokio teksto autorių verifikavimo metodai yra tikslesni, o jo padirbimas yra sudėtingas, tačiau statinio teksto duomenys yra prieinami lengviau, nes jų skaitmenizavimui nereikia jokių specialių prietaisų, o rašymas ant popieriaus yra natūralesnis ir priimtinesnis daugelyje sričių.

Abiejų teksto tipų atžvilgiu, teksto autoriaus verifikavimas išlieka sudėtingas uždavinys, dėl didelės duomenų variacijos klasės viduje, kai to paties žmogaus raštas, skirtingų rašymų metu, gali skirtis gana stipriai.

1.2. Rašysenos atpažinimo skirstymas pagal priklausomumą nuo teksto

Rašysenos atpažinimo metodus galima klasifikuoti į dvi grupes: nuo teksto priklausomus ir nepriklausomus [AS14]. Nuo teksto priklausomose sistemose ir mokymo, ir prognozavimo metu, autoriaus identifikavimui ar verifikavimui naudojamas tas pats tekstas, kuris dažniausiai būna sudarytas iš pavienių žodžių arba simbolių. Nuo teksto nepriklausomose sistemose, mokymui ir autoriaus identifikavimui, gali būti pateikiamas bet koks tekstas.

Nuo teksto priklausomas rašysenos atpažinimas sutinkamas retai, o viena pagrindinių tokių

sistemų pritaikymo sričių yra parašų atpažinimas, nes tas pats parašas (tekstas) yra naudojamas ir mokymui, ir identifikavimui. Nuo teksto nepriklausomas rašysenos atpažinimas pritaikomas praktikoje ten, kur vienodo teksto surinkimas yra nepriimtinas, pavyzdžiui kriminalistikoje, identifikuojant nusikaltimo įtariamąjį.

1.3. Rašysenos atpažinimo skirstymas pagal priklausomumą nuo autoriaus

Rašysenos atpažinimo sistemos, taip pat gali būti klasifikuojamos į dvi grupes pagal jų paskirtį: nuo rašančiojo priklausomas ir nepriklausomas [PJB+08]. Nuo rašančiojo priklausomų sistemų paskirtis yra identifikuoti rašantįjį, t.y. nustatyti jo tapatybę, o nepriklausomos sistemos atlieka rašančiojo verifikavimą, t.y. nustato ar rašantysis yra tas, kuo jis tvirtina esąs.

Nuo rašančiojo priklausomose sistemose kiekvienam teksto autoriui yra reikalingas atskiras modelis. Teksto autoriaus identifikavimo metu yra atliekamas vienas su daug (*angl. one-to-many*) palyginimas. Tokių sistemų trūkumas yra tai, kad kiekvienam naujam autoriui reikalingas naujas modelis, o jo apmokymui reikalingas didelis kiekis duomenų. Nuo rašančiojo nepriklausomose sistemose yra naudojamas vienas bendras modelis visiems autoriams, o verifikavimo metu yra atliekamas palyginimas vienas su vienu (*angl. one-to-one*). Abu sistemų tipai yra panašūs tuo, kad grąžina tam tikrą panašumo matą tarp teksto pavyzdžių, tačiau identifikavimo metu yra atliekama daugiaklasė klasifikacija, nustatanti autoriaus tapatybę, o verifikavimo metu binarinė klasifikacija, patvirtinanti arba paneigianti autoriaus tapatybę.

1.4. Rašysenos duomenų bazės

Rašysenos atpažinimo darbuose, dominuoja trys pagrindinės abėcėlės: lotynų, kinų ir arabų, todėl plačiausiai naudojamos, ranka rašyto teksto duomenų bazės, yra būtent šiomis abėcėlėmis. Lotynų kalbos abėcėlės daugiausiai sudarytos iš anglu kalbos teksto. Keletas populiarių rašysenos duomenų bazių pateikiama toliau.

IAM [MB02] statinio teksto duomenų bazė yra sudaryta iš 13353 anglų kalbos teksto eilučių, kurias parašė 657 autoriai. Ši duomenų bazė yra prieinama ne tik taip teksto eilutės, bet ir kaip pavieniai žodžiai, sakiniai ir pilni skenuoti teksto puslapiai. IAM taip pat turi ir dinaminio teksto duomenų bazę [LB05], sudarytą iš 13049 eilučių, parašytų 21 autoriaus.

RIMES [GCB+09] dinaminio teksto anglų kalbos duomenų bazė sudaryta iš 12723 puslapių laiškų, iš daugiau nei 1300 autorių. Duomenų bazė surinkta paprašius savanorių parašyti laišką, vienu iš devynių aprašytų scenarijų susijusių su verslo ir klientų santykiais. RIMES duomenų rinkinys naudotas neviename ICDAR ir ICFHR konkurse.

CASIA [LYW+11] dinaminio ir statinio teksto duomenų bazės yra vieni didžiausių kinų rašto duomenų rinkinių. Tiek dinaminis, tiek statinis tekstas šioje duomenų bazėje gautas vienu metu, naudojant specialų rašiklį, leidžiantį rašyti ant popieriaus, kartu išgaunant ir dinaminius rašymo duomenis. Ir statinio, ir dinaminio teksto duomenų rinkiniai pasirašyti 1020 autorių ir yra sudaryti iš 3.9 milijono simbolių iš 7356 skirtingų klasių.

1.5. Rašysenos atpažinimo darbų apžvalga

Rašysenos atpažinimas yra gana plačiai tyrinėta tema per pastaruosius keletą dešimtmečių. Vieni pirmųjų darbų pasirodė dar dvidešimto amžiaus dešimtajame dešimtmetyje [PL89]. Per šį laiką rašysenos atpažinimo metodai pasistūmėjo į priekį gana stipriai, o keletas jų apžvelgiama šiame skyriuje.

Marti ir kt. [MMB01] pristatė lotyniškos abėcėlės rašytojų identifikavimo sistemą. Autorių rašysenos atpažinimui šiame darbe išskirti dvylika požymių: plotis, aukštis, pasvirimas ir kiti rašysenos ypatumai. Jie panaudojo du klasifikatorius: k artimiausių kaimynų (*angl. k nearest neighbors, kNN*) ir tiesioginio sklidimo neuroninį tinklą (*angl. feed forward neural network*). Eksperimente panaudoti 100 puslapių teksto, parašyto 20 skirtingų autorių iš IAM duomenų bazės. Pasiekti 87.8% ir 90.7% tikslumai, atitinkamai su k artimiausių kaimynų ir neuroninio tinklo modeliais.

Chapran [CHA06] pristatė pilnai veikiančią, dinaminio teksto, rašytojų identifikavimo sistemą. Šiame darbe panaudotas naujas ypatybių atrinkimo algoritmas, paremtas panašumo koeficientais, aprašančiais kintamumo laipsnį, tarp to paties ir tarp skirtingų autorių rašyto teksto. Darbe aprašyti keletas klasifikatorių: minimalaus atstumo (*angl. minimum distance classifier*), Bajeso (*angl. Bayes classifier*) ir abiejų šių klasifikatorių kombinacijos. Darbe naudotas pačių autorių surinktas duomenų rinkinys, naudojant skaitmeninę planšetę ir spaudimui jautrų rašiklį. Duomenų rinkinys sudarytas iš 45 autorių, iš kurių kiekvienas parašė po 125 pavienius žodžius. Modeliui

pavyko teisingai identifikuoti 95% autorių.

Kitame 2006 metų darbe [SB06] pristatyta statinio, nuo teksto nepriklausoma, rašytojų identifikavimo sistema paremta Gauso mišinių modeliais (*angl. Gaussian mixture models, GMM*). Šiame darbe aprašytą modelį jie palygino su kita, su savo pačių pristatyta sistema [SB07], paremta paslėptaisiais Markovo modeliais (*angl. hidden Markov models, HMM*). Abiejuose darbuose buvo panaudotas tas pats duomenų rinkinys iš IAM duomenų bazės, sudarytas iš 100 autorių ir 4103 eilučių. GMM modeliui pavyko pasiekti didesnį 98.46% tikslumą, palyginus su HMM modeliu, kurio tikslumas siekė 97.03%.

Amaral ir kt. [AFB12] pristatė grafometrija paremtą būdą, rašytojų atpažinimui. Autorių rašysenos ypatybės, panaudotos šiame darbe, paremtos grafometrijos mokslu: reliatyvus teksto išdėstymas popieriaus lape ir atskirų žodžių aukščių santykiai. Iš išskirtųjų ypatybių sukonstruojamas 64 sveikųjų skaičių vektorius, kuris naudotas kaip įvestis atraminių vektorių klasifikatoriui (*angl. support vector machine, SVM*). Kaip duomenų rinkinys, panaudota „Brazilian forensic letter“ duomenų bazė, sudaryta iš 534 skirtingų autorių, statinių laiškų, kurių kiekvienas autorius parašė po 3. Šiame darbe aprašytas modelis pasiekė 80% tikslumą.

2013 metų darbe [BOJ+13], pristatyti autorių identifikavimo ir verifikavimo modeliai paremti tekstūrų operatoriais. Darbe teksto ypatybių išskyrimui panaudoti lokalūs binariniai šablonai (*angl. local binary patterns, LBP*) ir lokalus fazių kvantavimas (*angl. local phase quantization, LPQ*). Šiais metodais išskirtos ypatybės, perduotos atraminių vektorių klasifikatoriui. Darbe panaudoti „Brazilian forensic letter“ ir IAM duomenų rinkiniai. Rezultatai, gauti su LPQ ypatybių išskyrimo metodu, yra geresni nei LBP, ir rašytojų verifikavimo užduotyje su abiem duomenų rinkiniais siekė didesnį nei 99% tikslumą.

Chu ir Srihari [CS14] pristatytame darbe, sukurta rašytojų identifikavimo sistema, naudojant giliuosius neuroninius tinklus (*angl. deep neural network, DNN*). Šiame darbe aprašytas modelis, autorių identifikavimui, naudojo iš žodžių išskiriamus rašysenos požymius. Modelių mokymui ir testavimui, naudotas pačių šio darbo autorių surinktas 330 pradinių klasių mokinių, statinio teksto, duomenų rinkinys. Viena iš šio darbo išvadų, tai jog tokiam modeliui didelę įtaką turi pašaliniai veiksniai, tokie kaip pats žodis ar žodžio ilgis, o sėkmingam tokio modelio veikimui reikalingas kuo didesnis mokymo duomenų kiekis.

2015 metų darbe [YJL15], pristatyta DeepWriterID, nuo teksto nepriklausoma, dinaminio teksto, rašytojų identifikavimo sistema, paremta giliaisiais konvoliuciniais neuroniniais tinklais (*angl. deep convolutional neural network, CNN*). Šiame darbe jie pristatė naują duomenų papildymo metodą, pavadinimu DropSegment. Šis metodas pašalina atsitiktinius segmentus iš teksto, taip pagerinant modelio apibendrinamumą ir apsaugant nuo persimokymo. Eksperimente naudoti du „National Laboratory of Pattern Recognition“ rašysenos duomenų rinkiniai, kinų ir anglų kalbomis. Kinų kalbos rašysenos rinkinys sudarytas iš 187 autorių, o anglų kalbos rinkinys sudarytas iš 134 autorių. Darbe aprašytam modeliui pavyko pasiekti 95.72% ir 98.51% rašytojų identifikavimo tikslumą, atitinkamai, su kinų ir anglų kalbos duomenų rinkiniais.

2. Neuroninių tinklų apžvalga

Neuroniniai tinklai yra viena iš dirbtinio intelekto sričių. Neuroniniai tinklai yra įkvėpti žmogaus nervų sistemos. Jie yra sudaryti iš vieno ir daugiau sluoksnių neuronų, kurie, kaip ir žmogaus organizme, yra aktyvuojami esant tam tikroms sąlygoms. Šiame skyriuje apžvelgiami neuroninių tinklų tipai, juose sutinkami sluoksniai, neuronų aktyvacijos ir tikslo funkcijos, neuroninių tinklų mokymas.

2.1. Neuroniniai tinklų tipai

Įvairiuose darbuose galima rasti daug aprašytų neuroninių tinklų tipų [Sch14], o juos klasifikuoti galima pagal įvairius bruožus, pvz.: struktūrą, sluoksnių skaičių, duomenų sklidimą ir t.t.. Keletas populiariausių neuroninių tinklų yra: tiesioginio sklidimo neuroniniai tinklai, rekurentiniai neuroniniai tinklai, konvoliuciniai neuroniniai tinklai.

Tiesioginio sklidimo neuroniniai tinklai (*angl. feed-foward neural network*) – yra viena paprasčiausių neuroninių tinklų formų, kurioje duomenys keliauja tik viena kryptimi, iš ankstesnių sluoksnių į tolimesnius [BG94]. Tokie tinklai visada turi įvesties ir išvesties sluoksnius, o priklausomai nuo sudėtingumo gali turėti ir vieną ar daugiau paslėptųjų sluoksnių.

Rekurentiniai neuroniniai tinklai (*angl. recurrent neural network, RNN*) – priešingai nei tiesioginio sklidimo neuroniniai tinklai, turi jungčių ne tik į vėlesnius sluoksnius, bet ir į prieš tai ėjusius sluoksnius, taip sukuriant grįžtamąjį (rekurentinį) ryšį tarp sluoksnių [She18]. Šie tinklai plačiai taikomi teksto ir kalbos apdorojimo uždaviniuose.

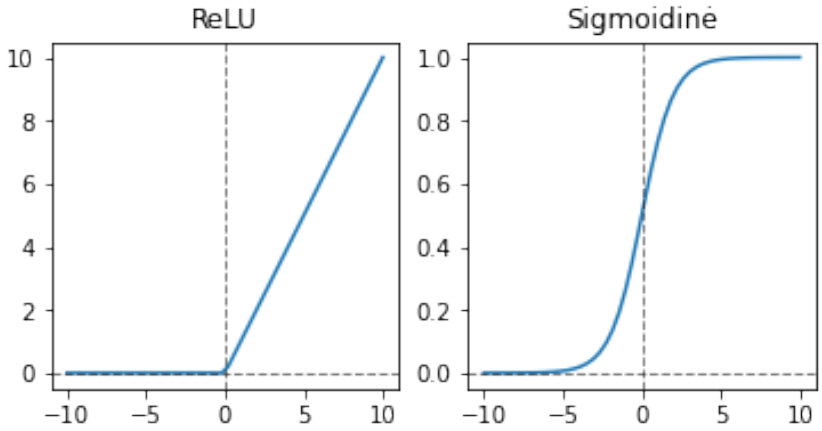
Konvoliuciniai neuroniniai tinklai (*angl. convolutional neural networks, CNN*) – vieni populiariausių šiuo metu naudojamų neuroninių tinklų, kurių pagrindinis pritaikymas yra vaizdų apdorojimo uždaviniuose. Susidomėjimą šiais tinklais paskatino AlexNet tinklo [KSH12] laimėjimas ILSVRC-2012 konkurse. Pagrindinė šių tinklų sudedamoji dalis yra konvoliuciniai sluoksniai, išskiriantys svarbius bruožus iš įvesties, naudodami filtro branduolius (*angl. kernel*) ir konvoliucijos operacijas. Vienas iš tokių tinklų privalumų, tai jog filtrų parametrus, reikalingus bruožų išskyrimui jis išmoksta pats.

Siamo tinklai (*angl. siamese network*) yra sudaryti iš dviejų ar daugiau identiškų subtinklų, pirmą kartą pristatyti spręsti parašų verifikavimo problemai [BBB+93]. Visi subtinklai turi iden-

tiškas architektūras, parametrus ir neuronų svorius. Pabaigoje šie tinklai dažniausiai yra sujungti atstumo funkcija, leidžiančia apskaičiuoti tam tikrą abiejų įvesčių panašumo matą. Mokymo metu, neuronų svorių atnaujinimas vykdomas identiškai visuose subtinkluose, todėl juose svoriai atnaujinami vienodai. Šie tinklai dažniausiai naudojami verifikavimo užduotyse, kur reikia nustatyti ar pora įvesčių priklauso tai pačiai, ar skirtingoms klasėms.

2.2. Aktyvacijos funkcijos

Neuroninio tinklo aktyvacijos funkcijos nusako, kaip įvesčių suma yra transformuojama į išvestį. Daugelis aktyvacijos funkcijų turi ribotą išvesties intervalą. Aktyvacijos funkcijos gali būti tiesinės arba ne tiesinės, priklausomai nuo transformacijos tipo. Daugeliu atveju visi paslėptieji modelio sluoksniai naudoja tokią pačią aktyvacijos funkciją, o išvesties sluoksnis dažniausiai naudoja kitokią aktyvacijos funkciją, priklausomai nuo modelio sprendžiamo uždavinio tipo [Sza20]. Pačios paprasčiausios aktyvacijos funkcijos yra linijinė (*angl. linear*), vienetinė (*angl. identity*) ar dvejetainė žingsninė (*angl. binary step-function*). Keletas sudėtingesnių, dažnai sutinkamų aktyvacijos funkcijų apžvelgtos plačiau.



1 pav. Aktyvacijos funkcijų palyginimas ReLU (kairėje), sigmoidinė (dešinėje)

Sigmoidinė (*angl. sigmoid*) aktyvacijos funkcija, dar vadinama logistine funkcija – ne tiesinė funkcija, sutinkama tiek paslėptuosiuose, tiek išvesties sluoksniuose. Šios funkcijos išvestis yra intervale tarp 0 ir 1. Kuo didesnė yra šios funkcijos įvestis, tuo arčiau išvestis bus prie 1, o kuo įvestis yra mažesnė, tuo išvestis bus arčiau 0. Išvesties sluoksnyje ši funkcija naudojama sprendžiant binarinio klasifikavimo uždavinius, kur jos išvestis gali būti interpretuojama kaip tikimybė, kad įvestis

priklauso tam tikrai klasei. Vienas iš šios funkcijos trūkumų, tai kad ji susiduria su nykstančio gradiento problema (*angl. vanihsing gradient*) [Hoc98], kai funkcijos reikšmės pasiekia kreivės galus, gradientas tampa labai mažas, o neurono mokymasis tampa lėtas arba visai sustoja.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

Išlyginta tiesinė aktyvacijos funkcija (*angl. rectified linear activation function, ReLU*), viena dažniausiai sutinkamų funkcijų naudojamų paslėptuosiuose sluoksniuose. Šios funkcijos išvestis apskaičiuojama, kaip maksimumas tarp 0 ir įvesties reikšmės, t.y. visoms neigiamoms reikšmėms bus grąžinamas 0, o visoms ne neigiamoms grąžinama įvesties reikšmė (2 formulė). Šią funkciją yra paprasta įgyvendinti, ir ji yra atspari nykstančių gradientų problemai. ReLU ir sigmoidinės aktyvacijų funkcijų grafikai pateikiami 1 paveikslėlyje.

$$f(x) = \max(0, x) \tag{2}$$

Softmax aktyvacijos funkcija – ne tiesinė funkcija, naudojama daugiaklasio klasifikavimo uždaviniuose išvesties sluoksnyje. Kaip įvestį Softmax aktyvacijos funkcija gauna realiųjų skaičių vektorių, kurio dydis yra klasifikuojamų klasių skaičius, o šios funkcijos išvestis yra, tokio paties dydžio vektorius, kurio reikšmių suma yra 1 (3 formulė), o jo reikšmės gali būti interpretuojamos kaip tikimybės, kad įvestis priklauso tam tikrai klasei.

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}} \tag{3}$$

2.3. Tikslo funkcijos

Tikslo funkcijos yra skirtos įvertinti, kaip gerai modeliui pavyksta atlikti spėjimus su tam tikrais duomenimis. Jeigu daugelis modelio spėjimų yra neteisingi, arba tolimi tikrosioms reikšmės, tikslo funkcijos rezultatas bus didelis, jeigu modelio spėjimai yra teisingi, rezultatas bus mažesnis. Keičiant ar mokant modelį, tikslo funkcijos rezultatas padeda įvertinti, ar pakeitimai ir tolimesnis mokymas yra naudingi. Keletas dažnai sutinkamų tikslo funkcijų aprašytos šiame skyriuje.

Vidutinė kvadratinė paklaida (*angl. mean square erorr, MSE*) – viena paprasčiausių ir po-

puliariausių tikslo funkcijų, sutinkama regresijos uždaviniuose. Šios tikslo funkcijos rezultatas yra vidurkis visų prognozuotų ir tikrų reikšmių skirtumų kvadratų (4 formulė). Vidutinės kvadratinės paklaidos rezultatas visada yra teigiamas. Skirtumų kvadratas reiškia, kad didelės modelio spėjimo klaidos, turės didesnę įtaką tikslo funkcijos rezultatui, nei mažos.

$$L = \frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{4}$$

Binarinė kryžminė entropija (*angl. binary cross-entropy*) – tikslo funkcija naudojama binarinės klasifikacijos problemose, kur rezultatas yra 0 arba 1. Kryžminė entropija apskaičiuoja vidutinį skirtumą tarp tikrosios ir spėtosios reikšmių tikimybių (5 formulė).

$$L = -\frac{1}{N} \cdot \sum_{i=1}^{N} y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i)) \tag{5}$$

Reta kategorinė kryžminė entropija (*angl. sparse categorical cross-entropy*) – tikslo funkcija skirta daugiaklasio klasifikavimo uždaviniui, kur rezultatas yra tarp 0 ir n, kur kiekviena klasė turi unikalią reikšmę. Kaip ir binarinės kryžminės entropijos atveju, ši funkcija apskaičiuoja vidutinį skirtumą tarp tikrosios ir spėtosios reikšmių tikimybių, tačiau kategorinės entropijos atveju, tai paskaičiuojama visoms klasėms (6 formulė). Jeigu pavyzdžiai gali priklausyti keletui klasių, naudojama kategorinė kryžminė entropija (*angl. categorical cross-entropy*), o etikečių sudarymui, neuroniniams tinklams su šia tinklo funkcija, naudojamas fiktyvus kodavimas (*angl. one-hot encoding*), kur kategorinės etiketės, pakeičiamos skaičių vektoriais.

$$L = -\sum_{i=1}^{N} y_i \cdot \log(p(y_i)) \tag{6}$$

Kontrastinė tikslo funkcija (*angl. contrastive loss*) – sutinkama Siamo architektūros tinkluose. Ši tikslo funkcija yra paremta ne spėjimo klaida, o atstumu tarp dviejų įvesčių, ir bando užtikrinti, kad atstumas tarp panašių įvesčių yra mažesnis, nei atstumas tarp skirtingų įvesčių. Kontrastinė tikslo funkcija visada skaičiuojama su įvesčių poromis. Ši funkcija aprašyta 7 formulėje, kur y – tikroji reikšmė, 0 arba 1, d – gautasis atstumas tarp įvesčių, m – numatytasis atstumas tarp įvesčių.

$$L = y \cdot d^2 + (1 - y) \cdot \max(m - d, 0)^2 \tag{7}$$

2.4. Optimizavimo algoritmai

Neuroninio tinko mokymo metu, yra keičiami neuronų svoriai, bandant minimizuoti tikslo funkcijos rezultatą, taip leidžiant modeliui atlikti kuo tikslesnius spėjimus. Mokymo metu pirmiausiai yra apmokomi išėjimo sluoksniai, klaidą perduodant paslėptiesiems sluoksniams, o toks mokymo būdas vadinamas atgalinės sklaidos (*angl. backpropagation*) metodu [RHW86]. Svorių keitimo užduočiai naudojami optimizavimo algoritmai. Praktikoje dažniausiai naudojami gradiento nusileidimo (*angl. gradient descent*) optimizavimo algoritmai. Gradiento nusileidimo algoritmai leidžia minimizuoti tikslo funkcijos rezultatą, keičiant modelio parametrus, priešinga gradientui kryptimi. Mokymo žingsnis (*angl. learning rate*) yra vienas iš šios funkcijos parametrų, nusakantis kokio dydžio žingsniais, bus keičiamas gradientas, taip užtikrinant, kad algoritmas konverguotų į minimumą.

Gradiento nusileidimo algoritmus galima išskirti į tris grupes: paketų gradiento nusileidimo (*angl. batch gradient descent*), stochastinio gradiento nusileidimo (*angl. stochastic gradient descent*) ir mini paketų gradiento nusileidimo (*angl. mini-batch gradient descent*), priklausomai nuo to, kiek duomenų yra apdorojama prieš kiekvieną modelio parametrų atnaujinimą [RHW86]. Paketų gradiento nusileidimo metodas, apskaičiuoja gradientą, visam naudojamam duomenų rinkiniui. Dėl šios priežasties šis metodas yra gana lėtas ir netinkamas dideliems duomenų rinkiniams, kurie netelpa sistemos atmintyje. Stochastinis gradiento nusileidimas, atnaujina neuronų svorius su kiekvienu mokymo pavyzdžiu. Mini paketų nusileidimo gradientas yra tarpinis variantas tarp ankščiau minėtųjų algoritmų. Šis algoritmas parametrus atnaujina, kas tam tikrą nurodytą skaičių, n, mokymo pavyzdžių.

Viena iš problemų, su kuria susiduriama apmokant modelius, yra mokymo žingsnio pasirinkimas. Per mažas mokymo žingsnis, sulėtina tikslo funkcijos konvergavimą, o per didelis mokymo žingsnis gali sukelti tikslo funkcijos svyravimą aplink minimumą ar net divergavimą. Šiai problemai spręsti sukurti gradiento nusileidimo algoritmai su kintamais mokymo žingsniais, pvz.: Adagrad [DHS11], Adadelta [Zei12], RMSprop [TH12] ar Adam [KB14].

2.5. Neuroninių tinklų sluoksniai

Neuroniniai tinklai yra sudaryti iš sluoksnių, kurių kiekvieną sudaro tam tikras skaičius neuronų. Visi neuroniniai tinklai turi įvesties, paslėptuosius ir išvesties sluoksnius. Tinklas sudarytas tik iš įvesties ir išvesties sluoksnių vadinamas vienasluoksniu, nes įvesties sluoksnis neatlieka jokių operacijų. Dažniausiai sutinkami sluoksniai neuroniniuose tinkluose yra: pilnai sujungti, konvoliuciniai, sutelkimo, normalizacijos ar išmetimo sluoksniai.

Pilnai sujungtas sluoksnis (*angl. fully connected (dense) layer*) – yra sluoksnis, kurio visi neuronai yra sujungti su visais ankstesniojo sluoksnio neuronais. Šis sluoksnis atlieka įvesties vektoriaus ir branduolio (*angl. kernel*) daugybą. Branduolio parametrai yra keičiami, apmokant modelį, atgalinės sklaidos pagalba. Pilnai sujungto sluoksnio išvestis yra n dydžio vektorius, kur n yra vienas iš šio sluoksnio parametrų, o išvesčiai dažnai pritaikoma viena iš aktyvacijos funkcijų. Šis sluoksnis yra sutinkamas daugelyje neuroninių tinklų modelių, tiek kaip paslėptasis, tiek kaip išvesties sluoksnis. Neuroniniuose tinkluose, kuriuose įvestis yra ne vienmatė, prieš šį sluoksnį, dažniausiai eina ištiesinimo (*angl. flatten*) sluoksnis, kuris keleto dimensijų įvestį konvertuoja į vienos dimensijos.

Konvoliucinis sluoksnis (*angl. convolutional layerl*) – yra konvoliucinio neuroninio tinklo pagrindas, pritaikytas darbui su dvimačiais duomenimis, dažniausiai nuotraukomis, bet jo panaudojimas galimas ir vienmačiams ar trimačiams duomenims. Šiame sluoksnyje yra atliekamos konvoliucijos operacijos. Dvimačių duomenų atveju, konvoliucijos operacija yra daugyba, tarp įvesties ir dvimačio masyvo, vadinamo filtro branduoliu (*angl. kernel*). Filtro branduolio dimensijos yra mažesnės už įvesties duomenų dimensijas, ir jis yra pritaikomas įvesčiai, slenkant filtrą, leidžiant jam persidengti, per visą įvestį, kas tam tikrą apibrėžtą žingsnį (*angl. stride*). Konvoliucijos operacijos išvestis yra vienas skaičius, kuriam papildomai pritaikoma aktyvacijos funkcija. Pritaikius konvoliucijas visai dvimačiai duomenų įvesčiai gaunamas dvimatis savybių žemėlapis (*angl. feature map*). Atliekant konvoliucijos operaciją įvesties kraštuose, filtras gali išeiti už įvesties ribų. Tokiu atveju atliekama papildymo (*angl. padding*) operacija, kuri elementams išeinantiems už įvestis ribų, priskiria tam tikrą reikšmę, dažniausiai 0. Daugeliu atveju konvoliucinis sluoksnis būna sudarytas iš keleto filtrų, išskiriančių skirtingas savybes.

Sutelkimo sluoksnis (*angl. pooling layer*) – sluoksnis sutinkamas konvoliuciniuose neuro-

niniuose tinkluose, dažniausiai einantis po konvoliucinio sluoksnio. Sutelkimo sluoksnis leidžia sumažinti savybių žemėlapį, taip išskiriant tik svarbiausias savybes. Sutelkimo sluoksnyje, kaip ir konvoliuciniame, taip pat yra naudojamas filtras, kuris slenka per visą įvestį. Šio filtro atliekama operacija gali būti dviejų tipų: sutelkimo imant maksimalią reikšmę (*angl. max pooling*) arba sutelkimo vidurkinant (*angl. average pooling*). Sutelkimo imant maksimalią reikšmę filtro išvestis yra didžiausia įvesties reikšmė, o sutelkimo vidurkinant filtro išvestis yra visų įvesties reikšmių vidurkis. Sutelkimo sluoksnis neturi apmokomų parametrų ir nedalyvauja modelio mokymo procese.

Normalizacijos sluoksniai (*angl. normalization layer*) – sluoksniai skirti įvesčių normalizavimui atlikti, taip stabilizuojant mokymo procesą. Vienas iš normalizacijos sluoksnių pavyzdžių yra paketų normalizacijos sluoksnis (*angl. batch normalization*) [IS15]. Paketų normalizacijos sluoksnis pritaiko transformacijas įvesčiai taip, kad išvesties vidurkis yra artimas 0, o standartinis nuokrypis artimas 1. Šio sluoksnio veikimas skiriasi modelio mokymo ir prognozavimo metu. Mokymo metu, duomenys bus normalizuojami dabartinio įvesties paketo vidurkiu ir standartiniu nuokrypiu. Prognozių atlikimo metu, duomenys normalizuojami, naudojant vidurkį ir standartinį nuokrypį, paketų kuriuos modelis matė mokymo metu.

Išmetimo sluoksnis (*angl. dropout layer*) – sluoksnis, kuris nurodytu dažniu atsitiktines įvesties reikšmes, pakeičia į 0 reikšmes, taip padedant sumažinti persimokymo (*angl. overfitting*) galimybę [HSK+12]. Įvestys, kurių reikšmės nėra nustatomos į 0, yra padidinamos $1/(1-dažnis)$, taip kad visų įvesčių suma lieka nepakitusi. Išmetimo sluoksnis yra pritaikomas tik mokymo proceso metu.

3. Neuroninis tinklas rašysenos verifikavimui

Šiame skyriuje aprašyta praktinė darbo dalis, kurios metu sukurti du Siamo architektūros neuroniniai tinklai, skirti statinio teksto, nuo rašančiojo ir teksto nepriklausomam, rašysenos atpažinimui. Šių tinklų architektūros, įkvėptos DeepWriter [XQ16] ir SigNet [DDT+17] tinklų.

3.1. Vieno srauto neuroninis tinklas

Pirmasis darbe apmokytas neuroninis tinklas paremtas parašų verifikavimo užduočiai pritaikytu SigNet tinklu. Tinklas kaip įvestį gauna 155×220 dydžio nuotrauką. Pirmajame konvoliuciniame sluoksnyje pritaikomi $96, 11 \times 11$ dydžio filtrai. Antrasis yra rinkinio normalizacijos sluoksnis, o po jo seka maksimalaus telkimo ir papildymo nuliais sluoksniai. Toliau eina antrasis konvoliucinis sluoksnis sudarytas iš $256, 5 \times 5$ dydžių filtrų, o po jo normalizacijos ir maksimalaus telkimo sluoksniai. Po šių sluoksnių seka atsitiktinio praretinimo sluoksnis su 0.3 išmetimo dažniu ir papildymo nuliais sluoksnis. Trečiajame konvoliuciniame sluoksnyje pritaikomi $384, 3 \times 3$ dydžio filtrai, o po jo nulinio papildymo sluoksnis. Ketvirtajame, paskutiniajame, konvoliuciniame sluoksnyje pritaikomi $256, 3 \times 3$ dydžio filtrai, o po šio sluoksnio eina maksimalaus telkimo ir atsitiktinio praretinimo sluoksniai. Toliau eina ištiesinimo sluoksnis ir pilnai sujungtas sluoksnis, su 1024 išvesties dydžiu. Po pilnai sujungto sluoksnio pritaikoma atsitiktinio praretinimo sluoksnis su 0.5 išmetimo dažniu. Paskutinysis sluoksnis yra pilnai sujungtas sluoksnis su 128 išvesties dydžiu. Visuose konvoliuciniuose ir pilnai sujungtuose sluoksniuose panaudota ReLU aktyvacijos funkcija.

$$d(y_1, y_2) = \sqrt{\sum_{i=1}^{K} (y_{1i} - y_{2i})^2} \tag{8}$$

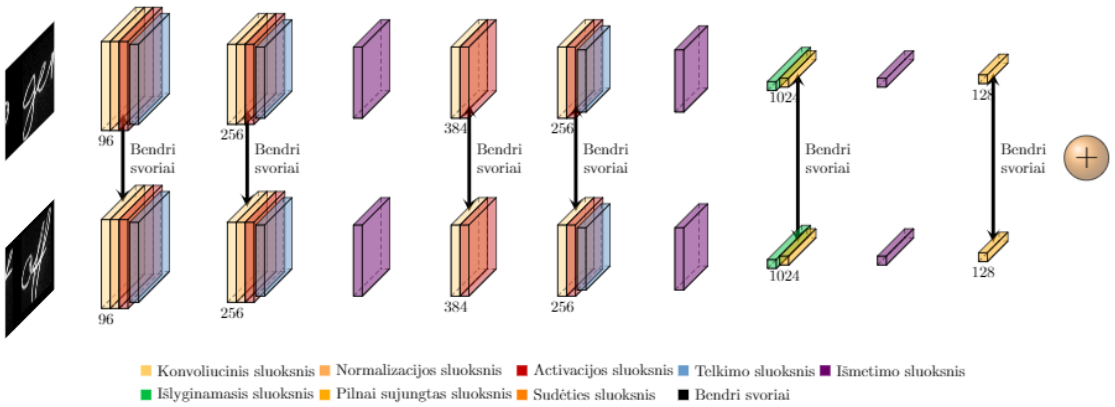
Neuroninis tinklas yra Siamo architektūros, todėl jis sudarytas iš dviejų identiškų, aukščiau aprašytų tinklų. Kaip įvestį šis tinklas gaus teksto atvaizdų porą, o po paskutiniojo sluoksnio šie tinklai sujungiami, ir yra apskaičiuojamas atstumas, tarp abiejų tinklų išvesčių. Kaip atstumo matas, pasirinktas euklidinis atstumas (8 formulė), o atstumo slenksčio (*angl. threshold*) įvertis, teksto fragmentų porų klasifikavimui kaip to paties ar skirtingų autorių, pasirinktas 0.5, t.y. poros, tarp kurių atstumas yra mažesnis nei 0.5, bus klasifikuojamos, kaip to paties autoriaus, o poros, tarp

kurių euklidinis atstumas didesnis ar lygus 0.5, bus klasifikuojamos kaip skirtingų autorių.

3.2. Dviejų kanalų neuroninis tinklas

Antrasis tinklas yra panašus į pirmąjį, tačiau pagrindinė šio tinklo idėja, gauta iš DeepWriter tinklo, skirto autorių identifikavimo užduočiai, yra perduoti tinklui ne vieną, o dvi gretimas, vieno autoriaus, teksto fragmentų iškarpas.

Visi šio neuroninio tinklo sluoksniai ir jų parametrai yra identiški pirmajam tinklui. Pagrindinis šio tinklo skirtumas, tai kad abiejų kanalų svoriai yra bendri visuose konvoliuciniuose ir pilnai sujungtuose sluoksniuose (2 paveikslėlis). Paskutiniųjų, kiekvieno kanalo, pilnai sujungtų sluoksnių išvestis yra perduodama sudėties sluoksniui, kurio išvestis yra sudėtos abiejų įvesties vektorių reikšmės.



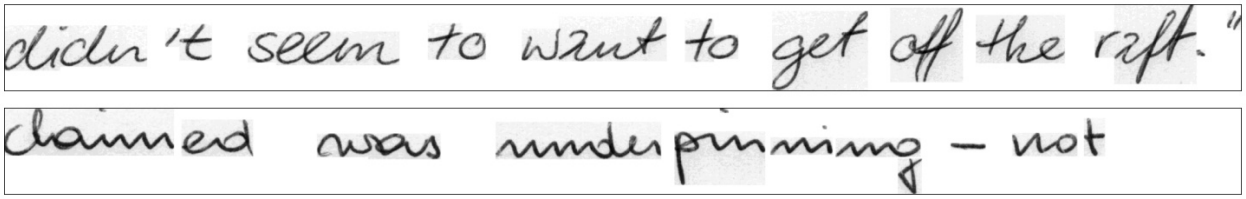
2 pav. Dviejų srautų tinklas, rodyklės tarp tinklų žymi bendrus svorius

Kaip ir pirmasis, šis dviejų kanalų tinklas yra Siamo architektūros, todėl kiekvieną Siamo šaką sudarys, po du lygiagrečius tinklus, o kaip įvestį gaus, ne vieną, o dvi poras teksto fragmentų, kur kiekvienoje poroje yra gretimi teksto fragmentai ir visa įvestis bus sudaryta iš 4 nuotraukų.

3.3. Duomenų rinkinys

Neuroninių tinklų mokymui ir įvertinimui pasirinktas „IAM Handwriting Database 3.0“ [MB02] statinio teksto duomenų rinkinys, sudarytas iš 657 autorių ranka parašyto teksto. Šiame darbe panaudota dalis šio duomenų rinkinio, sudaryta iš 13353 teksto eilučių nuotraukų (3

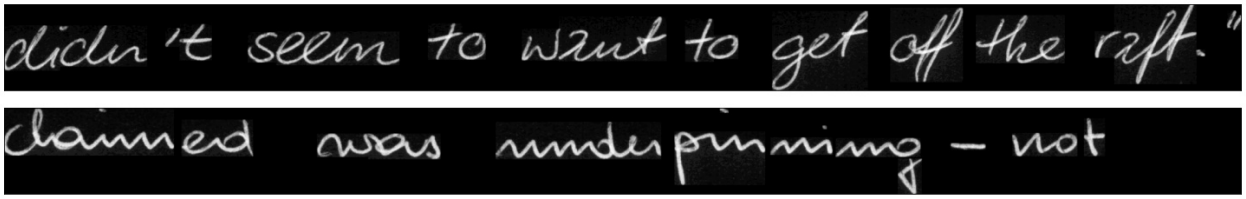
paveikslėlis). Kiekvienas autorius parašė tarp 2 ir 582 eilučių teksto, o rašymui naudotos įvairių tipų ir storių rašymo priemonės.



3 pav. 536 (viršuje) ir 020 (apačioje) autorių teksto eilučių pavyzdžiai

3.4. Duomenų apdorojimas

Prieš perduodant nuotraukas neuroniniam tinklui, jos buvo apdorotos keliais būdais. Duomenų rinkinyje teksto eilučių dydžiai varijavo nuo 38 iki 342 pikselių aukščio ir nuo 100 iki 2270 pikselių pločio. Visų nuotraukų dydis buvo pakeistas į fiksuotą 155 pikselių aukštį, o plotį pakeičiant taip, kad būtų išsaugomos pradinės teksto eilutės proporcijos.



4 pav. 536 (viršuje) ir 020 (apačioje) autorių apdorotų teksto eilučių pavyzdžiai

Pasirinktajame duomenų rinkinyje teksto eilučių nuotraukos skenuotos ir jose nebuvo pašalinio triukšmo, todėl, nuotraukų apdorojimo metu, nebuvo pritaikytas joks triukšmo pašalinimo metodas. Taip pat nebuvo pritaikytas joks binarizacijos algoritmas, tam kad nebūtų prarasti reikšmingi požymiai.

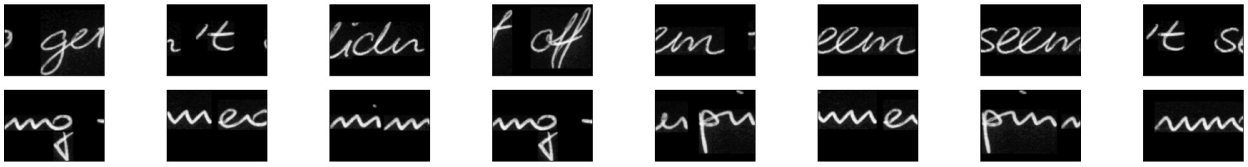
Pakeitus dydį visos nuotraukos invertuojamos, taip kad balti fono pikseliai pakeičiami juodais, 0 reikšmę turinčiais pikseliais, o teksto pikseliai į baltus, ne 0 reikšmės pikselius (4 paveikslėlis). Paskutiniame žingsnyje, pikselių reikšmės normalizuojamos, padalinant jas iš 255, taip kad visos pikselių reikšmės būtų tarp 0 ir 1.

3.5. Duomenų rinkinio išskirstymas

Neuroninio tinklo mokymui ir testavimui, visi duomenys padalinti į tris dalis: mokymo, validacijos ir testavimo duomenų rinkinius. Iš pradinių duomenų, atsitiktinai atrinkta po 60 autorių, kurių visos teksto nuotraukos priskirtos validacijos ir testavimo duomenų rinkiniams. Visų likusių 537 autorių nuotraukos panaudotos mokymo duomenų rinkiniui. Mokymo rinkinys buvo sudarytas iš 11113, validacijos iš 1082, testavimo iš 1158 teksto eilučių.

3.6. Teksto fragmentų karpymas ir porų generavimas

Neuroninio tinklo mokymui ir testavimui, generuotos dviejų tipų teksto nuotraukų poros: su abiem to paties autoriaus teksto nuotraukomis ir su skirtingų autorių tekstų nuotraukomis. Poros generuotos gyvai, mokant tinklą, t.y. kiekvienos tinklo mokymo iteracijos metu, buvo sugeneruojamos vis naujos poros iš susidarytųjų duomenų rinkinių. Generuojant nuotraukų poras, atsitiktinai parinkti autoriai ir jų teksto eilutės, iš kurių atsitiktinėje vietoje iškirptas 220 pikselių pločio teksto fragmentas, sudarant 155×220 dydžio fragmentą (5 paveikslėlis). Toks porų generavimo metodas leidžia sukurti, didelį skaičių, nepasikartojančių, nuotraukų porų.



5 pav. 536 (viršuje) ir 020 (apačioje) autorių, atsitiktinai iškirpti teksto eilučių fragmentai

3.7. Neuroninių tinklų mokymas

Paprastesniam modelių palyginimui, abu neuroniniai tinklai buvo apmokyti vienodai. Optimizavimo algoritmas pasirinktas RMSprop su šiais parametrais: mokymo žingsniu 1×10^{-4}, $\rho = 0.9$, $epsilon = 10^{-8}$. Kontrastinė funkcija pasirinkta kaip tikslo funkcija. Naudotas 128 nuotraukų porų paketo dydis (*angl. batch size*). Mokymas atliktas 500 epochų, su ankstyvu mokymo stabdymu (*angl. early stopping*), kai validacijos rinkinio tikslo funkcijos rezultatas, nebemažėja 10 epochų. Visi mokymo parametrai taip pat pateikiami 1 lentelėje.

Parametras	Reikšmė
Mokymo žingsnis	1×10^{-4}
ρ	0.9
ϵ	1×10^{-8}
Paketo dydis	128
Epochų skaičius	100

1 lentelė. Modelių mokymo parametrai

Tinklų architektūrai įgyvendinti ir eksperimentams atlikti panaudota Python programavimo kalbos, mašininio mokymo, Keras aplikacijų programavimo sąsaja ir Tensforflow platforma. Tinklų mokymas atliktas Google Colab aplinkoje, naudojant grafikos apdorojimo įrenginius (GPU).

Vieno kanalo neuroninio tinklo mokymas vyko 101 epochą ir užtruko apie 5 valandas. Dviejų kanalų neuroninio tinklo mokymas vyko 110 epochų ir užtruko apie 6 valandas.

3.8. Neuroninių tinklų mokymo rezultatai

Abiejų, vieno ir dviejų kanalų, tinklų testavimas vykdytas vienodai. Testavimo metu, buvo atlikti du eksperimentai, iš testavimui atsidėtų 60 autorių tekstų nuotraukų.

3.8.1. Rezultatų įverčių matai

Modelių įvertinimu pasirinkti 5 matai: tikslumas (*angl. accuracy*), preciziškumas (*angl. precision*), atšaukimas (*angl. recall*), vienodos klaidos tikimybė (*angl. equal error rate, EER*) ir plotas po ROC kreive (*angl. area under receiver operating characteristic (ROC) curve, AUC*).

Visus šiuos įverčius galima apskaičiuoti naudojant klasifikavimo lentelę (*angl. confussion matrix*) (2 lentelė), kurioje yra apibrėžiami visi įmanomi modelio spėjimo rezultatai. Teisingai teigiamas (*angl. true positive, TP*) rezultatas, kai modelis teisingai atspėja teigiamą klasę, o teisingai neigiamas (*angl. true negative, TN*) rezultatas, kai modelis teisingai atspėja neigiamą klasę. Neteisingai teigiamas (*angl. false positive, FP*) rezultatas yra tada, kai modelis neteisingai prognozuoja teigiamą klasę, o neteisingai neigiamas (*angl. false negative, FN*) rezultatas, kai modelis neteisingai spėja neigiamą klasę.

	Teigiama reikšmė	Neigiama reikšmė
Teigiamas spėjimas	Teisingai teigiamas rezultatas (TP)	Neteisingai teigiamas rezultatas (FP)
Neigiamas spėjimas	Neteisingai neigiamas rezultatas (FN)	Teisingai neigiamas rezultatas (TN)

2 lentelė. Klasifikavimo lentelė

Tikslumo įvertis yra dydis, nusakantis kokią dalį spėjimų modelis atliko teisingai ir yra apibrėžiamas kaip visų teisingų spėjimų ir visų atliktų spėjimų santykis (9 formulė).

$$tikslumas = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

Preciziškumas nusako, kokia dalis teigiamų spėjimų, buvo teisinga, ir yra apibrėžiamas, kaip teisingai teigiamų ir teisingai teigiamų, bei neteisingai teigiamų spėjimų sumos santykis (10 formulė).

$$preciziškumas = \frac{TP}{TP + FP} \tag{10}$$

Atšaukimo įvertis nusako, kokia dalis teigiamų reikšmių, buvo identifikuota teisingai ir yra apibrėžiamas kaip teisingai teigiamų ir teisingai teigiamų, bei neteisingai neigiamų spėjimų sumos, santykis (11 formulė).

$$atšaukimas = \frac{TP}{TP + FN} \tag{11}$$

ROC kreivė, yra grafikas parodantis modelio veikimą su visomis klasifikavimo slenksčių reikšmėmis. Šios kreivės abscisių ašyje atvaizduojama neteisingai teigiamų spėjimų dalį, o ordinačių ašyje, teisingai teigiamų spėjimų dalį.

Teisingai teigiamų spėjimų dalis (*angl. true positive rate, TPR*) – atitinka atšaukimo įvertį ir yra apibrėžiama, kaip teisingai teigiamų ir teisingai teigiamų, bei neteisingai neigiamų spėjimų sumos santykis (12 formulė).

$$TPR = \frac{TP}{TP + FN} \tag{12}$$

Neteisingai teigiamų rezultatų dalis (*angl. false positive rate, FPR*), apibrėžiama kaip netei-

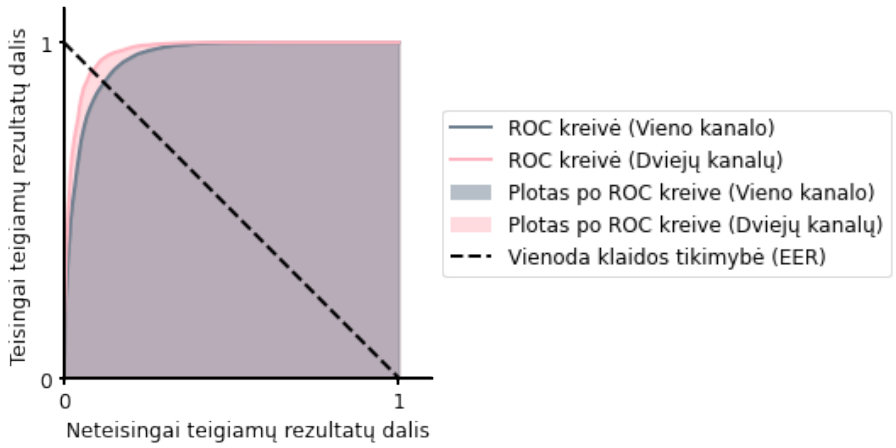
singai teigiamų ir neteisingai teigiamų, bei teisingai neigiamų spėjimų sumos santykis (13 formulė).

$$FPR = \frac{FP}{FP + TN} \tag{13}$$

Plotas po ROC kreive, yra dydis nusakantis plotą po visa ROC kreive, tarp koordinačių $(0, 0)$ ir $(1, 1)$. Šis dydis pateikia bendrą sistemos įvertį, su visais klasifikavimo slenksčiais.

3.8.2. Modelių su vienu teksto fragmentu

Pirmajame eksperimente iš testavimo rinkiniui atsidėtų 60 autorių, sugeneruotos 16384 nuotraukų poros, taikant tokį patį porų generavimo metodą, kaip ir modelio mokymo ir validacijos metu. Teksto fragmentų poros klasifikavimui, kaip to paties ar skirtingų autorių, kaip ir modelių mokymo atveju paliktas, tas pats, 0.5 atstumo slenksčio įvertis.



6 pav. ROC kreivės

Vieno kanalo modeliui pavyko pasiekti 88.12% tikslumą, su 83.96% preciziškumo ir 94.25% atšaukimo įverčiais. Šio tinklo vienoda klaidos tikimybė yra ties 12.00%, o plotas po ROC kreive 0.950.

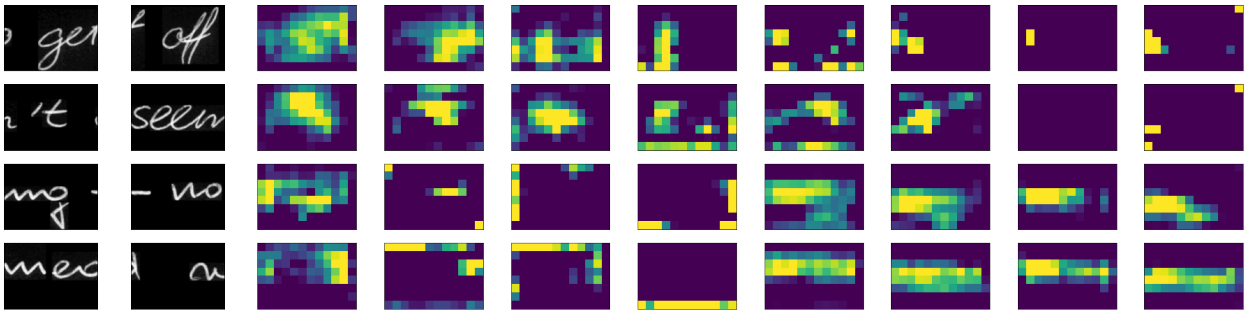
Dviejų kanalų modeliui pavyko pasiekti 91.74% tikslumą, su 89.29% preciziškumo ir 94.85% atšaukimo įverčiais. Šio tinklo vienoda klaidos tikimybė yra ties 8.53%, o plotas po ROC kreive 0.970.

Visi modelių palyginimo įverčiai pateikiami 3 lentelėje, o ROC kreivė pateikiama 6 paveikslėlyje.

Modelio įvertis	Vieno kanalo	Dviejų kanalų
Tikslumas (%)	88.12	91.74
Preciziškumas (%)	83.96	89.29
Atšaukimas (%)	94.25	94.85
Vienoda klaidos tikimybė (%)	12.00	8.53
Plotas po ROC kreive	0.950	0.970

3 lentelė. Modelių įverčiai

Dviejų kanalų tinklas pralenkia vieno kanalo tinklą visuose įverčiuose. Vienas iš pastebėjimų tai, kad nors abiejų modelių atšaukimo matas yra labai panašus, dviejų kanalų tinklo preciziškumas yra gerokai didesnis, nei vieno kanalo tinklo. Tai reiškia, kad abu modeliai, bandydami verifikuoti autorių, neteisingai atmes panašų procentą, to paties autoriaus teksto, tačiau dviejų kanalų tinklas sugeneruos mažesnį kiekį neteisingai teigiamų spėjimų. Šis dydis yra svarbus sprendžiant užduotis, kur kito žmogaus teigiamas verifikavimas, yra nepriimtinas, pavyzdžiui įleidžiant autorių į sistemą su jautriais duomenimis.



7 pav. 536 (2 viršuje) ir 020 (2 apačioje) autorių, savybių žemėlapiai

7 paveikslėlyje pavaizduoti 8 skirtingų filtro branduolių, paskutiniojo konvoliucinio sluoksnio savybių žemėlapiai, su didelėmis aktyvacijos reikšmėmis. Galima pastebėti, kad to paties autoriaus teksto fragmentai yra artimesni vienas kitam, nei skirtingų.

3.8.3. Modelių rezultatai su keliais teksto fragmentais

Pirmasis eksperimentas sekė tokią pat eigą, kaip ir mokymo, bei validacijos metu, kur prognozė atliekama pagal vieną teksto iškarpų nuotraukų porą. Tokiame eksperimente yra neišnaudojama didelė dalis turimų duomenų, nes viena teksto iškarpa sudaro tik nedidelę dalį viso autoriaus

parašyto teksto.

Antrojo eksperimento metu, buvo sudaromos atsitiktinės autorių poros, tačiau kiekvienai autorių porai buvo sudaromos daugiau nei viena teksto fragmentų $(2, 4, 8, 16)$ pora. Abiems modeliams atlikti spėjimą buvo perduodama tik po vieną porą, tačiau pilnas spėjimas buvo atliekamas tik apskaičiavus visų, vienai autorių porai sugeneruotų, teksto fragmentų porų atstumų aritmetinį vidurkį. Visų teksto porų klasifikavimui, kaip to paties ar skirtingų autorių, naudotas 0.5 atstumo slenksčio įvertis.

Vieno kanalo tinklui su 16 teksto fragmentų porų pavyko pasiekti 91.77% tikslumą, su 85.97% preciziškumo ir 100.00% atšaukimo įverčiais. Šio tinklo vienoda klaidos tikimybė, su 16 teksto fragmento porų, yra 5.48%, o plotas po ROC kreive 0.977.

Dviejų kanalų tinklui su 16 teksto fragmentų porų pavyko pasiekti 95.05% tikslumą, su 91.14% preciziškumo ir 100.00% atšaukimo įverčiais. Šio tinklo vienoda klaidos tikimybė, su 16 teksto fragmento porų, yra 3.34%, o plotas po ROC kreive 0.988.

Galima pastebėti, kad kaip ir pirmojo eksperimento metu, dviejų kanalų modelis, visuose įverčiuose, su visais nuotraukų porų skaičiais, pralenkia vieno kanalo modelį. Visi abiejų modelių įverčiai su skirtingais nuotraukų fragmentų skaičiais pateikiami 4 lentelėje.

Modelio įvertis	**VK1**	VK2	VK4	VK8	VK16	**DK1**	DK2	DK4	DK8	DK16
Tikslumas (%)	**88.12**	90.52	90.94	93.02	91.77	**91.74**	92.86	94.74	94.48	95.05
Preciziškumas (%)	**83.96**	85.76	85.41	87.91	85.97	**89.29**	89.02	91.04	90.27	91.14
Atšaukimas (%)	**94.25**	97.40	99.27	99.90	100.00	**94.85**	97.92	99.37	99.79	100.00
EER (%)	**12.00**	8.34	7.74	6.06	5.48	**8.53**	7.05	4.84	4.17	3.34
AUC	**0.950**	0.965	0.966	0.979	0.977	**0.970**	0.970	0.980	0.982	0.988

4 lentelė. Modelių įverčiai su keliomis teksto iškarpomis

Gautuose rezultatuose galima pastebėti keletą bendrų tendencijų. Pirmoji tendencija yra ta, kad didinant teksto fragmentų skaičių, pagal, kuriuos yra sprendžiama ar tai to paties ar skirtingų autorių rašysena, abiejų modelių atveju tikslumas gerėja. Tai yra svarbu jeigu turimas tikrojo ir tiriamojo autoriaus teksto kiekis yra nemažas, tokiu atveju galima atlikti užtikrintesnius spėjimus. Ši tendencija neturi didelės įtakos, jei turimų duomenų kiekis yra mažas.

Įdomu tai, kad abiejų modelių atvejais, su 16 teksto fragmentų nuotraukų porų, pasiektas

100% atšaukimo įvertis, tai reiškia, kad abu modeliai teisingai identifikavo visas to paties autoriaus teksto fragmentų poras. Tačiau, kaip ir pirmojo eksperimento metu, dviejų kanalų modelio preciziškumo įvertis buvo didesnis už vieno kanalo tinklą. Taip pat galima pastebėti, kad nors atšaukimo įvertis, atliekant spėjimus su 16 teksto fragmentų porų, pakilo daugiau nei 5%, palyginus su 1 teksto fragmento pora, preciziškumo įvertis pakilo tik per 2% abiejų modelių atveju. Priklausomai nuo sprendžiamos problemos tipo, galima būtų padidinti slenksčio dydį, taip sumažinant teisingai teigiamų spėjimų skaičių, bet padidinant neteisingai neigiamų spėjimų skaičių.

Išvados

Autorių atpažinimas – tai viena iš kompiuterinės vaizdų analizės sričių, leidžiančių identifikuoti ar verifikuoti teksto autorių, pagal jo rašyseną, dažnai sutinkama biometrinėse sistemose ar kriminalistikoje, tiriant dokumentų autentiškumą.

Šiame darbe susipažinta su teksto autorių atpažinimo užduotimi ir joje taikomais metodais. Statinio, nuo teksto nepriklausomo, autorių verifikavimo užduočiai spręsti, buvo realizuoti du, vieno ir dviejų kanalų, Siamo architektūros konvoliuciniai neuroniniai tinklai, paremti SigNet ir DeepWriter tinklų architektūromis. Šių tinklų mokymui ir testavimui panaudotas „IAM handwriting database“ duomenų rinkinys, kaip įvestį tinklams perduodant atsitiktinius to paties arba skirtingų autorių, teksto fragmentų poras.

Apmokius neuroninius tinklus, geresni rezultatai gauti su dviejų kanalų tinklu, kuriam pavyko pasiekti 91.74% autorių verifikavimo tikslumą ir 8.53% vienodos klaidos tikimybės įvertį. Pabandžius spėjimą apie autorių atlikti iš daugiau nei vienos teksto fragmentų poros, gauti dar geresni rezultatai. Su dviejų kanalų tinklu pasiektas 95.05% tikslumas ir 3.34% vienoda klaidos tikimybė, atliekant spėjimą apskaičiavus 16 nuotraukų porų atstumo vidurkį.

Tolimesni tyrimai galėtų apimti neuroninių tinklų struktūros tobulinimą, dar geresnių rezultatų gavimui. Taip pat, norint šį modelį paversti pilnai funkcionuojančiu produktu, reikėtų įgyvendinti metodą, leidžiantį iš pilno dokumento nuotraukos išskirti atskiras teksto eilutes.

Literatūra

[AFB12] Aline Maria MM Amaral, Cinthia OA Freitas ir Flávio Bortolozzi. The graphometry applied to writer identification. *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, p. 1. The Steering Committee of The World Congress in Computer Science, Computer …, 2012.

[AS14] Ahmed Ahmed ir Ghazali Sulong. Arabic writer identification: a review of literature. *Journal of Theoretical and Applied Information Technology*, 69:474–484, 2014-11.

[BBB+93] Jane Bromley, James Bentz, Leon Bottou, Isabelle Guyon, Yann Lecun, Cliff Moore, Eduard Sackinger ir Rookpak Shah. Signature verification using a ”siamese” time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence*, 7:25, 1993-08.

[BG94] G. Bebis ir M. Georgiopoulos. Feed-forward neural networks. *IEEE Potentials*, 13(4):27–31, 1994.

[BOJ+13] D. Bertolini, L.S. Oliveira, E. Justino ir R. Sabourin. Texture-based descriptors for writer identification and verification. *Expert Systems with Applications*, 40(6):2069–2080, 2013. ISSN: 0957-4174.

[CHA06] JOULIA CHAPRAN. Biometric writer identification: feature analysis and classification. *International Journal of Pattern Recognition and Artificial Intelligence*, 20(04):483–503, 2006.

[CS14] Jun Chu ir Sargur Srihari. Writer identification using a deep neural network. *Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing*, ICVGIP ’14, Bangalore, India. Association for Computing Machinery, 2014. ISBN: 9781450330619.

[DDT+17] Sounak Dey, Anjan Dutta, Juan Ignacio Toledo, Suman K. Ghosh, Josep Lladós ir Umapada Pal. Signet: convolutional siamese network for writer independent offline signature verification. *CoRR*, abs/1707.02131, 2017.

[DHS11] John Duchi, Elad Hazan ir Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. 12(null):2121–2159, 2011-07. ISSN: 1532-4435.

[GCB+09] Emmanuèle Grosicki, Matthieu Carré, Jean-Marie Brodin ir Edouard Geoffrois. Results of the rimes evaluation campaign for handwritten mail processing. *2009 10th International Conference on Document Analysis and Recognition*, p. 941–945, 2009.

[Hoc98] Sepp Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6:107–116, 1998-04.

[HSK+12] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever ir Ruslan Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012.

[YJL15] Weixin Yang, Lianwen Jin ir Manfei Liu. Deepwriterid: an end-to-end online text-independent writer identification system, 2015.

[IOO91] S. IMPEDOVO, L. OTTAVIANO ir S. OCCHINEGRO. Optical character recognition — a survey. *International Journal of Pattern Recognition and Artificial Intelligence*, 05(01n02):1–24, 1991.

[IS15] Sergey Ioffe ir Christian Szegedy. Batch normalization: accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015.

[KB14] Diederik Kingma ir Jimmy Ba. Adam: a method for stochastic optimization. *International Conference on Learning Representations*, 2014-12.

[KSH12] Alex Krizhevsky, Ilya Sutskever ir Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[LB05] Marcus Liwicki ir H. Bunke. Iam-ondb - an on-line english sentence database acquired from handwritten text on a whiteboard. Tom. 2005, 956–961 Vol. 2, 2005-10.

[LYW+11] Cheng-Lin Liu, Fei Yin, Da-Han Wang ir Qiu-Feng Wang. Casia online and offline chinese handwriting databases. *2011 International Conference on Document Analysis and Recognition*, p. 37–41, 2011.

[LSB+06] Marcus Liwicki, Andreas Schlapbach, Horst Bunke, Samy Bengio, Johnny Mariéthoz ir Jonas Richiardi. Writer identification for smart meeting room systems. Horst Bunke ir A. Lawrence Spitz, redaktoriai, *Document Analysis Systems VII*, p. 186–195, Berlin, Heidelberg. Springer Berlin Heidelberg, 2006. ISBN: 978-3-540-32157-6.

[MB02] Urs-Viktor Marti ir H. Bunke. The iam-database: an english sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, 5:39–46, 2002-11. DOI: 10.1007/s100320200071.

[MMB01] Urs-Viktor Marti, R. Messerli ir H. Bunke. Writer identification using text line based features. P. 101–105, 2001-02. ISBN: 0-7695-1263-1.

[PJB+08] Daniel Pavelec, Edson Justino, Leonardo Batista ir Luiz Soares de Oliveira. Author identification using writer-dependent and writer-independent strategies. P. 414–418, 2008-01.

[PL89] Réjean Plamondon ir Guy Lorette. Automatic signature verification and writer identification — the state of the art. *Pattern Recognition*, 22(2):107–131, 1989. ISSN: 0031-3203.

[Pre] Cambridge University Press. Handwriting. *Cambridge dictionary*. URL: https://dictionary.cambridge.org/dictionary/english/handwriting (tikrinta 2021-05-24).

[RHW86] D. E. Rumelhart, G. E. Hinton ir R. J. Williams. *Learning internal representations by error propagation*. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*. MIT Press, Cambridge, MA, USA, 1986, p. 318–362. ISBN: 026268053X.

[SB06] A. Schlapbach ir H. Bunke. Off-linewriter identification using gaussian mixture models. *18th International Conference on Pattern Recognition (ICPR'06)*, tom. 3, p. 992–995, 2006.

[SB07] Andreas Schlapbach ir Horst Bunke. A writer identification and verification system using hmm based recognizers. *Pattern Anal. Appl.*, 10:33–43, 2007-02.

[Sch14] Jürgen Schmidhuber. Deep learning in neural networks: an overview. *CoRR*, abs/1404.7828, 2014.

[SCV09] Imran Siddiqi, Florence Cloppet ir Nicole Vincent. Contour based features for the classification of ancient manuscripts. *Conference of the International Graphonomics Society*, p. 226–229, 2009.

[She18] Alex Sherstinsky. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *CoRR*, abs/1808.03314, 2018.

[Sza20] Tomasz Szandala. Review and comparison of commonly used activation functions for deep neural networks. *CoRR*, abs/2010.09458, 2020.

[TH12] T. Tieleman ir G. Hinton. Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning, 2012.

[XQ16] Linjie Xing ir Yu Qiao. Deepwriter: A multi-stream deep CNN for text-independent writer identification. *CoRR*, abs/1606.06472, 2016.

[Zei12] Matthew D. Zeiler. ADADELTA: an adaptive learning rate method. *CoRR*, abs/1212.5701, 2012.

Priedas nr. 1

Tyrimams naudotas kodas

Šio darbo tyrimai atlikti Google Colab aplinkoje, Python programavimo kalba, naudojant mašininio mokymo, Keras aplikacijų programavimo sąsaja ir Tensforflow platforma. Nuoroda į Git versijų kontrolės repozitoriją, su šio tyrimo Python „užrašine“ ir apmokytais modeliais:

`https://github.com/kedvinas/HandwritingVerification.git`