Učenje podobnosti v globokih nevronskih omrežjih za razpoznavanje obrazov

Žiga Stržinar, Klemen Grm, Vitomir Štruc

University of Ljubljana, Faculty of Electrical Engineering,
Tržaška cesta 25, 1000 Ljubljana, Slovenia
E-mail: zs4002@student.uni-lj.si, {klemen.grm, vitomir.struc}@fe.uni-lj.si

Abstract

Učenje podobnosti med pari vhodnih slik predstavlja enega najpopularnejših pristopov k razpoznavanju na področju globokega učenja. Pri tem pristopu globoko nevronsko omrežje na vhodu sprejme par slik (obrazov) in na izhodu vrne mero podobnosti med vhodnima slikama, ki jo je moč uporabiti za razpoznavanje. Izračun podobnosti je pri tem lahko v celoti udejanjen z globokim omrežjem, lahko pa se omrežje uporabi zgolj za izračun predstavitve vhodnega para slik, preslikava iz izračunane predstavitve v mero podobnosti pa se izvede z drugim, potencialno primernejšim modelom. V tem prispevku preizkusimo 5 različnih modelov za izvedbo preslikave med izračunano predstavitvijo in mero podobnosti, pri čemer za poizkuse uporabimo lastno nevronsko omrežje. Rezultati naših eksperimentov na problemu razpoznavanja obrazov kažejo na pomembnost izbire primernega modela, saj so razlike med uspešnostjo razpoznavanje od modela do modela precejšnje.

1 Uvod

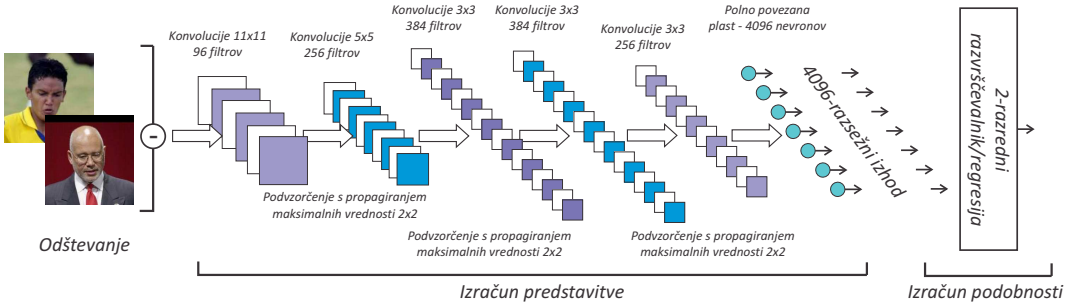
Globoka konvolucijska omrežja so v zadnjih letih postala nepogrešljivo orodje pri izgradnji učinkovitih sistemov računalniškega vida in umetne inteligence. Ključni dejavniki, ki so omogočili razcvet področja globokega učenja, je vse večja razpoložljivost računskih virov ter izjemne količine podatkov, ki omogočajo stabilno učenje globokih omrežij. Najnaprednejša omrežja so tako zagotovila do sedaj nevideno učinkovitost na različnih področjih kot so detekcija ali sledenje objektov v slikah in videu, razpoznavanje objektov oz. vizualnih kategorij v slikovnih podatkih, semantično opisovanje slik, ipd.

Težava, ki se pogosto pojavi pri izgradnji globokih nevronskih omrežij, pa je nezadostni obseg materiala, ki je na voljo za učenje. Izpostaviti namreč velja, da je tipično število parametrov, katerih vrednost je potrebno pri učenju sodobnih globokih omrežjih (namenjenim razpoznavanju obrazov) določiti, reda velikosti 10^8 in več. Zahtevani obseg učnih podatkov se zato meri v GB.

V [1] smo predstavili pristop, ki potrebo po velikih količinah učnih podatkov zaobide z učenjem globokega nevronskega omrežja na parih slik in s tem zagotovi, da je količina razpoložljivega učnega materiala kvadratična v številu razpoložljivih učnih slik. V tem prispevku uporabimo nadgradnjo pristopa iz [1], pri čemer vhoda v globoko nevronsko omrežje tokrat ne predstavlja par slik obrazov, ampak njuna razlika, kot je prikazano na Sliki 1. Z nevronskim omrežjem želimo določiti skalarno vrednost, ki ustreza podobnosti med slikama na vhodu omrežja. Omeniti velja, da se je razpoznavanje v prostoru razlik (angl. difference space) v preteklosti že izkazalo za izredno učinkovitega za namene verifikacije obrazov [2], [3]. Do neke mere je takšen pristop podoben tudi pristopom za učenje metrik (angl. metric learning) [4], [5], [6], ki prav tako poskušajo zgraditi računske modele, sposobne vračanja podobnosti med pari slik.

Čeprav lahko predlagani pristop v celoti udejanjimo z globokim nevronskim omrežjem, pa so se v literaturi kot učinkovitejši izkazali dvo stopenjski modeli, sestavljeni iz globokega omrežja, s katerim se iz para slik najprej določi značilke (oz. predstavitev para slik), izračunana predstavitev pa se nato s plitvim razvrščevalnikom (oz. regresijskim modelom) preslika v mere podobnosti. V tem prispevku se navežemo na te postopke in v prvem koraku zgradimo globoko nevronsko omrežje, ki na vhodu sprejme razliko dveh barvnih slik obrazov (velikosti 128×128 slikovnih elementov) ter na izhodu vrne 4096-razsežni vektor značilk (oz. 4096-razsežno predstavitev), ki opisuje vhodni par. Arhitekturo mreže, prikazabno na Sliki 1, pri tem povzamemo po AlexNet omrežju [7], ki je sestavljeno iz zaporedja konvolucijskih plasti ter plasti, namenjenih podvzorčenju. Velikost konvolucijskih filtrov pri tem iz ene v drugo konvolucijsko plast zmanjšujemo. Omrežje učimo preko softmax razvrščevalnika na podmnožici zbirke IJB-A [8], ki vsebuje 33200 parov slik (16600 parov slik z enako identiteto in 16600 parov slik z različno identiteto). Ko je omrežje naučeno, vrhnji softmax razvrščevalnik odstranimo ter dobljeni 4096-razsežni izhod globokega omrežja uporabimo za izračun predstavitve para slik.

Za izračunano predstavitev vhodnega para slik želimo nato poiskati model (tj., razvrščevalnik ali regresijski model), ki bo omogočil najučinkovitejšo preslikavo izračunane predstavitve v mero podobnosti (glej desno stran Slike 1). V ta namen v prispevku preizkusimo 5 različnih modelov: razvrščevalnik s podpornimi vektorji (angl. support vector machine oz. SVM), odločiteno drevo (angl. decision tree), naključne gozdove (angl. random forest), regresijo z metodo



Slika 1: Bločna shema uporabljene arhitekture globokega nevronskega omrežja.

najmanjših kvadratov (angl. least squares regression) in logistično regresijo (angl. logistic regression). Vse preizkušene modele uporabimo v regresijskem načinu, pri čemer za par slik, ki pripada isti identiteti, želimo izhodno vrednost 1 in za par slik, ki pripadata različnim identitetam, želimo izhodno vrednost 0.

Preostanek prispevka je razdeljen v tri razdelke. V drugem razdelku predstavimo modele, ki smo jih evalvirali v našem delu. V tretjem razdelku opišemo eksperimente na podatkovni zbirki IJB-A in podamo najpomembnejše ugotovitve. Prispevek zaključimo z nekaj opažanji v zadnjem, četrtem razdelku.

2 Učenje podobnosti

Nevronsko omrežje iz Slike 1 za par vhodnih slik vrne 4096-razsežno predstavitev, ki jo želimo s pomočjo izbranih modelov preslikati v mero podobnosti med vhodnima slikama. V tem razdelku predstavimo teoretično ozadje izbranih modelov, s katerimi smo v eksperimentanem delu izvedli preslikavo.

2.1 Razvrščevalnik s podpornimi vektorji

Razvrščevalnik s podpornimi vektorji (angl. support vector machine - SVM [9]) temelji na iskanju optimalne ločilne meje med dvema razredoma vzorcev v skladu s principom maksimalnega roba (angl. maximum margin). SVM algoritem stremi k čim večji razdalji med ločilno mejo in najbližjimi vzorci obeh razredov iz učne množice - tem vzorcem rečemo podporni vektorji (od tu tudi izvira ime metode).

Učne vzorce – točke v n dimenzionalnem prostoru značilk označimo z $\mathbf{x}_i \in \mathbb{R}^{n\times 1}$, kjer je $i = 1, ..., N$ in je N število vseh učnih vzorcev. Vsak učni vzorec pripada enemu od dveh razredov: $y_i \in \{-1, 1\}$, za $i = 1, ..., N$. Ob predpostavki linearno ločljivih podatkov lahko ločilno mejo izrazimo kot:

$$f(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b, \quad \text{kjer je } l \ll N, \qquad (1)$$

kjer so α_i skalarne uteži in b odmik, ki omogoča, da ločilna meja ne poteka nujno skozi koordinatno izhodišče n-razsežnega prostora. Nov vzorec $\mathbf{x}$ razvrstimo tako, da določimo predznak desne strani Enačbe (1).

Ločilna meja v Enačbi (1) je linearna, vendar lahko z uporabo različnih jeder $K(\mathbf{x}, \mathbf{x}_i)$ (Enačba (2)) dosežemo tudi nelinearne ločilne meje [3], [10], [11], [12]:

$$f(x) = \sum_i \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b. \qquad (2)$$

Učenje SVM razvrščevalnikov poteka preko namenskega optimizacijskega postopka, ki minimizira kriterijsko funkcijo z več odprtimi parametri. Med njimi omenimo parameter C, katerega vpliv v eksperimentalnem delu tudi preverjamo in določa razmerje med strmenjem algoritma k ločilni meji s čim-širšim robom (majhen C) in strmenjem k modelu s čim manj napačno razvrščenimi vzorci v učni množici (večji C) [12], [9]. Drug parameter, katerega vpliv prav tako ovrednotimo v nadaljevanju, ϵ, določa regularizacijo modela. Vpliva na gladkost ločilne meje in število podpornih vektorjev. S tem določa kompleksnost in generalizacijske sposobnosti modela [12], [9].

2.2 Odločitvena drevesa

Drug razvrščevalnik, ki nas v tem prispevku zanima so odločitvena drevesa. Ljudje sprejemamo številne delne odločitve, podobno pa delujejo tudi odločitvena drevesa. V vsakem vozlišču drevesa, se izvede delna odločitev na podlagi ene same značilke. Pri zveznih vrednostih značilk se v vsakem vozlišču drevo razveji na dve veji, ki ustrezata vrednostim pod in nad določenim pragom (prag in značilko, na podlagi katere se drevo veji, določimo z učenjem drevesa). Pri diskretnih značilkah se drevo veji na različne vrednosti značilke [11], [13], [12].

Za regresijsko izvedbo odločitvenega drevesa je tipično potrebno določiti:

- število značnilk, ki se v vsakem vozlišču drevesa preveri pri iskanju najboljše značilke za nov razcep; to število je lahko tudi enako številu značilk vzorca, kar seveda vpliva na računsko zahtevnost iskanja,
- največjo globino drevesa,
- najmanjše potrebno število vzorcev v vozlišču, da še pride do razcepa, in
- najmanjše potrebno število vzorcev v listu.

V eksperimentalnem delu smo za implementacijo odločitvenih dreves uporabili knjižnico *scikit learn*.

2.3 Naključni gozdovi

Tretji model, ki nas zanima v prispevku, so naključni gozdovi (angl. random forests).

Metoda naključnih gozdov ustvari večje število odločitvenih dreves, pri čemer je vsako drevo učeno

ločeno. V algoritem učenja je vključena še naključnost, saj se pri vsakem drevesu, v vsakem vozlišču, kot kandidate za razcep upošteva le naključno-izbrane podmnožice značilk [13], [12].

2.4 Linearna regresija

Linearna regresija, ki smo jo uporabili v prispevku, temelji na običajni metodi najmanjših kvadratov. Postopek išče takšno regresijsko matriko $\boldsymbol{\beta} \in \mathbb{R}^{n\times 1}$, ki bo regresor $\mathbf{x}_i \in \mathbb{R}^{n\times 1}$, tj., vzorec iz učne množice moči N, $i = 1, ..., N$, po Enačbi (3) preslikala v pravilno vrednost izhoda y_i. Pri tem je za probleme razvrščanja vrednost izhoda lahko -1 za prvi in 1 za drugi razred:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta}. \tag{3}$$

Ker imamo znano množico učnih vzorcev z znanimi oznakami y_i, imamo podan sistem enačb, ki ga lahko zapišemo matrično:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}, \tag{4}$$

kjer je vektor na levi $\mathbf{y} = [y_1, y_2, \ldots, y_N] \in \mathbb{R}^{N\times 1}$ in je $\mathbf{X} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_N^T] \in \mathbb{R}^{N\times n}$.

Sistem rešimo po Enačbi (5) in tako dobimo parametre modela [11], [12]:

$$\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{5}$$

2.5 Logistična regresija

Drugi primer regresijskega modela, ki ga obravnavamo v prispevku, je logistična regresija. Logistično regresijo lahko razumemo kot poseben primer linearne regresije z dodanim nivojem oz. dodano nelinearnostjo. Izhod razpoznavalnika z logistično regresijo je omejen na interval (0,1) in ga pogosto tolmačimo kot verjetnost pripadnosti razredu.

Logistična regresija uporablja funkcijo $g(z)$, ki izhod omeji na interval (0,1):

$$g(z) = \frac{e^z}{e^z + 1} = \frac{1}{1 + e^{-z}}. \tag{6}$$

Kot vhod v $g(z)$ lahko uporabimo izhod linearnega modela [14], [12] kot je definiran z Enačbo (4):

$$\mathbf{y} = g(\mathbf{X}\boldsymbol{\beta}) = \frac{1}{1 - e^{-\mathbf{X}\boldsymbol{\beta}}}. \tag{7}$$

3 Eksperimenti in rezultati

Uspešnost preslikave iz predstavitve para vhodnih slik, ki jo dobimo z globokim nevronskim omrežjem, v mero podobnosti med slikama smo preverili na problemu verifikacije obrazov in podatkovni zbirki IJB-A [8]. Podatkovna zbirka je bila zbrana v okviru DARPA projekta JANUS, katerega glavni cilj je povečati zanesljivost sistemov za razpoznavanje obrazov za en velikostni razred. Zbirka se uvršča med najzahtevnejše zbirke, ki so trenutno na voljo za področje razpoznavanja obrazov, saj je bila zbrana s spleta brez uporabe samodejnih detektorjev obrazov, ki bi nabor slik omejili na obraze, ki so jih detektorji sposobni zaznati. Za namene preizkušanja smo uporabili 16700 parov slik, od tega 8350 parov predstavlja ujemanja, preostalih 8350 parov pa so pari slik, ki pripadajo različnim identitetam. Za ponazoritev učinkovitosti razpoznavanja smo uporabili skalarne mere v obliki površine pod ROC krivuljami [15], ki jih v nadaljevanju označujemo kot AUC (angl. Area Under Curve).

Vse razvrščevalnike, ki smo jih na kratko predstavili v Razdelku 2, smo uporabili v regresijskem načinu. Z razvrščevalniki smo izvedli dve seriji poizkusov. V prvi seriji je bil naš namen preučiti vpliv velikosti učne množice na uspešnost različnih razvrščevalnikov. V drugi seriji pa smo se osredotočili na dva tipa razvrščevalnika in izvedli optimizacijo njunih parametrov. Primerjava uspešnosti vseh razvrščevalnikov je podana na koncu razdelka.

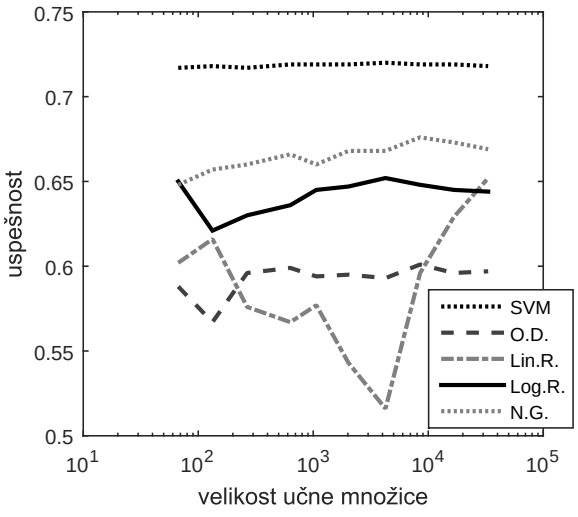
3.1 Vpliv velikosti učne množice

Vsakega od razvrščevalnikov smo preizkusili pri različnih velikostih učne množice, kot prikazuje Slika 2. Graf ponazarja uspešnost (tj. vrednosti AUC) pri različnih velikostih učne množice. Za preverjanje uspešnosti smo vedno uporabili vseh 16700 testnih vzorcev.

Opazimo, da na razvrščevalnik s podpornimi vektorji povečanje učne množice nima vpliva. Enako lahko trdimo za razpoznavalnik na osnovi odločitvenih dreves. Rezultati pri linearni regresiji močno variirajo. Pri logistični regresiji in razvrščevalniku po metodi naključnih gozdov pa je opazen rahel trend naraščanja uspešnosti z večanjem učne množice.

3.2 Vpliv hiperparametrov

V našem drugem eksperimentu smo želeli preveriti vpliv nekaterih parametrov razvrščevalnikov na njihovo uspešnost. Omejili smo se na SVM in odločitvena



Slika 2: Vpliv velikosti učne množice na uspešnost razvrščevalnikov. V legendi pomenijo: SVM (razvrščevalnik s podpornimi vektorji), O.D. (odločitvena drevesa), Lin. R (linearna regresija), Log. R. (logistična regresija), N.G. (naključni gozdovi). Uspešnost se meri v obliki vrednosti AUC.

drevesa, ki sta sodeč po rezultatih predstavljenih na Grafu 2 najboljši in najslabši razvrščevalnik.

V naših poizkusih smo se oredotočili SVM s polinomskim jedrom. Pri preizkušanju nas je zanimal vpliv parametrov C, ϵ ter stopnje polinomskega jedra. Parametra C in ϵ smo spreminjali v okolici privzetih vrednosti $C = 1$ in $\epsilon = 0,1$. Vpliv C smo raziskali v območju od 0,125 do 4, vpliv ϵ pa od 0,025 do 0,8. Stopnjo polinoma smo (seveda ob nespremenjenih – privzetih ostalih parametrih razvrščevalnika) spreminjali od 1 do 5. Uporabili smo 3320 učnih vzorev (10 % celotne učne množice) ter vseh 16700 testnih vzorcev. Vse eksperimente smo izvedli dvakrat, enkrat brez normiranja in enkrat z normiranjem vseh elementov predstavitve para slik med 0 in 1. Z eksperimentiranjem smo ugotovili, da spreminjanje parametrov C in ϵ nima bistvenega vpliva na uspešnost razvrščevalnika. Rezultati pri spreminjanju stopnje polinoma pa so prikazani v Tabeli 1.

Tabela 1: SVM s polinomskim jedrom: vpliv stopnje polinoma in normiranja vzorcev

stopnja polinoma	AUC	
	brez normiranja	z normiranjem
1	0,721	0,717
2	0,719	0,721
3	0,717	0,721
4	0,717	0,721
5	0,703	0,721

Pri optimizaciji odločitvenih dreves smo preverjali največjo dovoljeno globino odločitvenega drevesa, ki smo jo spreminjali od 1 do 14. Za prve eksperimente smo uporabili le 10 % vseh učnih vzorcev (3320 vzorcev oz. predstavitev). V rezultatih je bilo zaznati, da dobimo najboljše rezultate pri plitvejših drevesih. Eksperiment smo zato ponovili pri manjših globinah z večjo (polno) učno množico. Rezultate pri vseh 33200 učnih vzorcih prikazuje Tabela 2. Najboljše rezultate smo dobili pri globinah 3, 4 in 5. Za vrednotenje smo vedno uporabili vseh 16700 testnih vzorcev (predstavitev). Velja omeniti,

Tabela 2: Odločitveno drevo: Vpliv največje dovoljene globine

največja globina	AUC	*največja globina*	AUC
1	0.650	5	0.709
2	0.698	6	0.700
3	0.711	7	0.688
4	0.705	8	0.682

da privzeto največja dovoljena globina ni določena. Pri razpoznavalniku s privzetimi nastavitvami in uporabi 33200 učnih vzorcev smo dobili slabšo uspešnost (ROC AUC = 0,597) kot pri razpoznavalniku z omejitvijo dovoljene globine (globina 3: ROC AUC = 0,711).

3.3 Primerjava razvrščevalnikov

Na koncu podamo še primerjavo med vsemi preizkušenimi razvrščevalniki. Tabela 3 prikazuje najboljše rezultate za preizkušene razpoznavalnike, ki smo jih dosegli z optimizacijo odprtih parametrov.

Tabela 3: Primerjava vseh razvrščevalnikov. Najboljši rezultat smo dosegli z uporabo SVM razvrščevalnika.

Metoda	AUC
SVM	**0,721**
Odločitveno drevo	0,711
Linearna regresija	0,652
Logistična regresija	0,652
Naključni gozdovi	0,676

4 Zaključek

V prispevku smo preverili primernost 5 razvrščevalnikov za izvedbo preslikave med predstavitvijo para vhodnih slik obrazov, ki jo vrne globoko nevronsko omrežje, in podobnostjo med parom slik. Naši rezultati so pokazali, da se najbolje odreže razvrščevalnik s podpornimi vektorji (SVM), kar kaže na dejstvo, da je smiselno v prihodnje raziskati kriterijske funkcije za učenje globokih omrežij, ki temeljijo na principu maksimalnega roba in jih je mogoče integrirati v sam postopek izgradnje omrežja.

Literatura

[1] K. Grm, S. Dobrisek, and V. Struc, "Deep pair-wise similarity learning for face recognition," in *IWBF 2016*, 2016.

[2] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian face recognition," *Pattern Recognition*, vol. 33, pp. 1771–1782, 2000.

[3] P. J. Phillips, "Support vector machines applied to face recognition," in *NIPS*, 1998.

[4] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *CVPR 2005*, vol. 1, 2005, pp. 539–546.

[5] D. Yi, Z. Lei, S. Liao, and S. Li, "Deep metric learning for person re-identification," in *ICPR 2014*, 2014, pp. 34–39.

[6] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *IEEE CVPR*, 2015, pp. 815–823.

[7] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012, pp. 1097–1105.

[8] B. Klare, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge, and A. Jain, "Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a," in *CVPR 2015*, 2015, p. 9.

[9] V. Vapnik, *The nature of statistical learning theory*. Springer, New York, 1995.

[10] B. Heiselea, P. Ho, and T. Poggio, "Face recognition with support vector machines: global versus component-based approach," in *Proceedings Eighth IEEE International Conference on Computer Vision*, 2001.

[11] N. Pavešič, *Razpoznavanje vzorcev: uvod v analizo in razumevanje vidnih in slušnih signalov*. Založba FE in FRI, 2012.

[12] Scikit learn. [Online]. Available: http://scikit-learn.org/

[13] V. Ghosal, "Efficient face recognition system using random forests," Master's thesis, Indian Institute of Technology Kanpur, 2009.

[14] C. Zhou, L. Wang, W. Zhang, and X. Wei, "Face recognition based on pca and logistic regression analysis," *Elsevier*, 2014.

[15] R. Gajsek, V. Struc, S. Dobrisek, and F. Mihelic, "Emotion recognition using linear transformations in combination with video," in *Interspeech 2009*.