

Prepoznavanje emocija korišćenjem dubokih neuronskih mreža

Seminarski rad u okviru kursa
Računarska inteligencija
Matematički fakultet

Denis Aličić
a.denis96@gmail.com

Sažetak

U ovom radu je predstavljena duboka neuronska mreža koja rešava problem klasifikacije emocija na osnovu fotografije. Ulaz u neuronsku mrežu predstavlja vektor realnih brojeva dužine 128 (eng. *face embedding*) [9] koji preslikava sliku lica u euklidski vektorski prostor. Zatim je nad tim vektorima trenirana mreža. Za implementaciju modela je korišćeno okruženje Keras [4], koji olakšava upotrebu moćne biblioteke Tensorflow [1]. Kod koji prati rad je dostupan na <https://github.com/DenisAlicic/Face-emotions>.

1 Uvod

Prepoznavanje emocija predstavlja jedan od poznatih i još uvek u potpunosti nerešenih problema. Razvoj mašinskog učenja, konkretno oblasti konvolutivnih neuronskih mreža, doveo je do poboljšanja postojećih rešenja. Ovaj problem ima bitnu ulogu u oblasti međuljudskih odnosa [2], kao i odnosa čoveka i računara. Popularnosti ovog problema doprinelo je Kaggle takmičenje otvoreno 2013. godine, čiji je nagradni fond iznosio 500 dolara. U takmičenju je učestvovalo 56 timova. Većina postojećih rešenja koristi treniranje konvolutivnih mreža izgrađenih od nekoliko slojeva. S obzirom da sam pročitao rad [9] koji govori o preslikavanju lica u euklidski vektorski prostor dimenzije 128, kao i da svaki takav vektor (eng. *face embedding*) jedinstveno određuje lice,

odlučio sam da to iskoristim kao ulaz u neuronsku mrežu. Za treniranje i testiranje mreže je korišćen skup podataka „fer2013” [6].

2 Opis skupa podataka

Skup podataka „fer2013” [6] sadrži slike (eng. *grayscale*) slike lica veličine 48x48 piksela. Cilj je klasifikovati svako lice u jednu od 7 kategorija:

0. Ljutnja (eng. *angry*)
1. Gađenje (eng. *disgust*)
2. Strah (eng. *fear*)
3. Sreća (eng. *happy*)
4. Tuga (eng. *sad*)
5. Iznenadenje (eng. *surprise*)
6. Neutralno (eng. *neutral*)

Skup sadrži 28709 slika u delu za trening i 3589 slika u delu za testiranje. Skupovi za trening i test su dati u jednom csv fajlu koji sadrži piksele slike, kategoriju i informaciju o tome u kom skupu se nalazi dati red. Primer nekoliko slika iz skupa je dat na slici 1.

3 Preprocesiranje podataka

Od csv fajlova su generisane slike 48x48, a zatim je korišćena biblioteka „face recognition” koja implementira rad [9] da bi se dobili vektori dimenzije 128.



Slika 1: Primer slika iz skupa podataka

U ovom radu piše da ti vektori jedinstveno predstavljaju lice.

3.1 Više o vektorima

Mogu da se koriste za problem prepoznavanja lica, iz razloga što se slična lica mapiraju u slične vektore u dimenzionom prostoru u smislu euklidske bliskosti. Takođe je spomenuto da ovi vektori u sebi čuvaju informacije o facijalnoj ekspresiji. Upravo taj rad koji su napisali istraživači iz Guglove istraživačke laboratorije me inspirisao da pokušam da rešim ovaj problem koristeći taj pristup. Mapiranje lica u vektore se vrši iz dva dela. Prvi korak je nalaženje lica na fotografiji, a zatim se koristi konvolutivna mreža koja sadrži 17 slojeva da mapira lica u vektor. Ja nisam implementirao tu mrežu, već sam koristio gotovu implementaciju, iz razloga što hardverski resursi koje sam imao na raspolaganju to nisu mogli da urade. Proces treniranja takve mreže je izuzetno hardverski zahtevan.

4 Model

Model se sastoji od 7 slojeva. Prvi sloj u mreži normalizuje podatke, svodeći ih na normalnu (Gausovu) raspodelu. Ovaj pristup ubrzava proces učenja, tako što omogućava veći koeficijent učenja (eng. *learning rate*), praktično čini nebitnim za finalni model

izbor jezgra (eng. *kernel*) i poboljšava ukupne metrike modela [7]. Da bi se sprečilo preprilagođavanje modela, korišćena je L2 regularizacija [5]. Ona služi da ograniči prilagodljivost modela, tako da ga samim tim štiti od preprilagođenosti. Formula L2 regularizacije:

$$L' = L + \lambda \|\mathbf{w}\|_2 \quad (1)$$

gde je:

- L - funkcija gubitka
- $\|\mathbf{w}\|_2$ - L₂ norma vektora težina \mathbf{w}
- $L_2 = \sqrt{\sum_i \mathbf{w}_i^2}$
- λ - koeficijent regularizacije

Detalji o ostalih 6 slojeva su prikazani u tabeli 1.

Tabela 1: Model

Br. neurona	Aktivaciona fun.	Regul.
128	ReLU	L2
86	ReLU	L2
64	ReLU	-
32	Tanh	-
16	ReLU	L2
7	Softmax	-

Za aktivacionu funkciju poslednjeg sloja je izabrana funkcija „Softmax“:

$$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}} \quad (2)$$

kao jedna od najčešće korišćenih u problemima višeklasne klasifikacije.

4.1 Treniranje modela

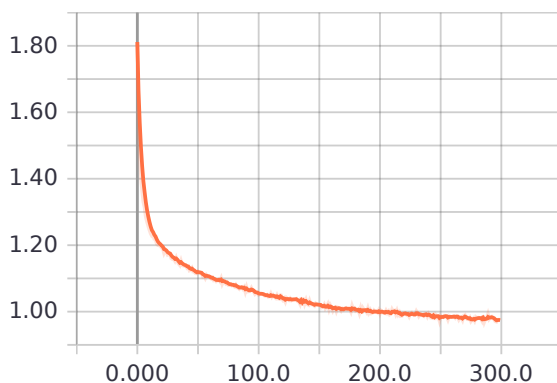
Model je treniran u 300 epoha. Funkcija gubitka koja je optimizovana pri treniranju je entropija:

$$E_{\text{entropy}} = - \sum_{c=1}^M y_{0,c} \log(p_{0,c}) \quad (3)$$

gde je:

- M - broj klasa,
- y - binarni indikator (0, 1) da li je klasa c ispravna klasa primera o ,
- p - predviđena verovatnoća da primerak o pripada klasi c .

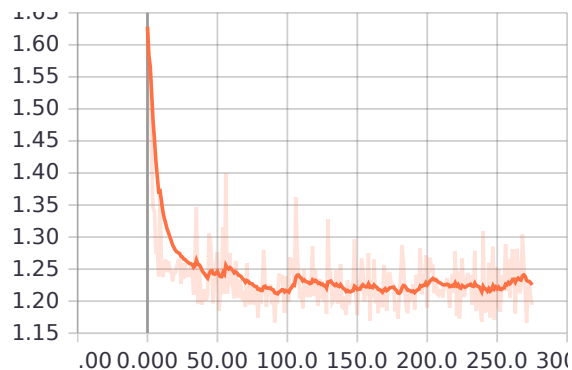
Optimizator korišćen prilikom treniranja je Adam opisan u radu [8]. Izbegavanje platoa prilikom minimizovanja funkcije je izvedeno redukovanjem koeficijenta (eng. *learning rate*) gradijentnog spusta. Opadanje funkcije gubitka na skupovima za testiranje i validiranje kroz vreme je prikazano na slikama 2 i 3.



Slika 2: Opadanje funkcije gubitka kroz vreme na skupu za treniranje

5 Rezultati

Metrika koja je posmatrana je tačnost (eng. *accuracy*). Tačnost je korišćena i u takmičenju, tako da je jednostavno porediti rezultate sa drugim radovima iz ove oblasti. Tačnost nije jedina metrika koja određuje uspešnost modela, tako da je vođeno računa i o tome da one klase za koje postoji malo podataka u trening i test skupu ne budu uvek promašene. Ovo je postignuto balansiranjem seta korišćenjem tehnike SMOTE (Synthetic Minority Over-sampling Technique) [3] koja kombinuje pristup slučajnog odabira već postojećih podataka i generisanja sintetičkih podataka korišćenjem metode „k najbližih suseda”.



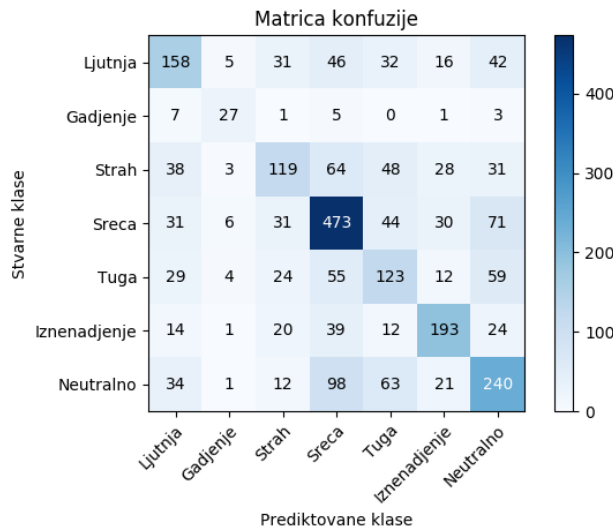
Slika 3: Opadanje funkcije gubitka kroz vreme na skupu za validiranje

Ukupna tačnost je 54% što bi na Kaggle takmičenju zauzelo 20. poziciju od mogućih 56. Pobednik takmičenja je postigao rezultat od 71%. Rezultate 10 prvoplasiranih je moguće videti u tabeli 2. Ovi rezultati možda deluju neubedljivo, ali anali-

Tabela 2: Rezultati takmičenja

Pozicija	Tačnost
1	71%
2	69,2%
3	68,8%
4	67,4%
5	65,2%
6	65%
7	64,4%
8	64,2%
9	63,1%
10	62,1%

zom skupa podataka se zaključi da je nekada dosta teško i čoveku odrediti koja je emocija na licu (npr. strah ili tuga, ljutnja ili tuga itd.) zbog toga što ljudi različito iskazuju svoje emocije. U prilog ovome govori i matrica konfuzije 4 iz koje se jasnije vidi u kojim situacijama model greši.



Slika 4: Matrica konfuzije

6 Zaključak

Rad predstavlja ispitivanje mogućnosti vektora lica (eng. *face embeddings*) za dati problem koji su prvenstveno namenjeni za prepoznavanje osoba (eng. *face recognition*). U budućnosti model je moguće dodatno poboljšati implementacijom mreže koja mapira lica u vektore i na te slojeve dodati arhitekturu mreže predstavljene u ovom radu i sve zajedno trenirati za specifičan problem uz eventualno dodavanje skupova podataka koji su dostupni.

Literatura

- [1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [2] Vinay Bettadapura. Face expression recognition and analysis: The state of the art. *CoRR*, abs/1203.6722, 2012.
- [3] Kevin W. Bowyer, Nitesh V. Chawla, Lawrence O. Hall, and W. Philip Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *CoRR*, abs/1106.1813, 2011.
- [4] François Chollet et al. Keras. <https://keras.io>, 2015.
- [5] Corinna Cortes, Mehryar Mohri, and Afshin Rostamizadeh. L2 regularization for learning kernels. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, pages 109–116, Arlington, Virginia, United States, 2009. AUAI Press.
- [6] Ian Goodfellow, Dumitru Erhan, Pierre-Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, Yingbo Zhou, Chetan Ramaiah, Fangxiang Feng, Ruifan Li, Xiaojie Wang, Dimitris Athanasakis, John Shawe-Taylor, Maxim Milakov, John Park, Radu Ionescu, Marius Popescu, Cristian Grozea, James Bergstra, Jingjing Xie, Lukasz Romaszko, Bing Xu, Zhang Chuang, and Yoshua Bengio. Challenges in representation learning: A report on three machine learning contests, 2013.
- [7] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015.
- [8] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014. cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.

- [9] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. *CoRR*, abs/1503.03832, 2015.