

GENERISANJE SLIKE NA OSNOVU VEŠTAČKE NEURONSKE MREŽE

Tomislav Dobrički SW-21/2014, Ognjen Vlajić SW-15/2014

MOTIVACIJA

Veštačke neuronske mreže se često korsite za klasifikaciju slika, ali sam način kojim neuronska mreža tokom treniranja nauči da prepozna je teško uvideti.

Cilj ovog softvera jeste da se razvije algoritam koji će na osnovu istrenirane neuronske mreže da generiše sliku za određene klase. Neuronska mreža bi trebala da generisanu sliku prepozna kao jednu od svojih klasa. U zavisnosti od kvaliteta generisane slike bi se ovim procesom potencijlano otkrili nedostaci neuronske mreže.

TEORETSKO REŠENJE

Veštačka neuronska mreža se sastoji iz više slojeva, a svaki sloj je skup više čvorova (neurona). Izlazna vrednost neurona jeste zbir njenih ulaznih vrednosti pomnoženih sa odgovarajućim težinama. Izlazna vrednost čvora a_4 sa slike 1:

$$a_4 = x_1 w_{41} + x_2 w_{42} + x_3 w_{43} + b$$

U formuli slovo b predstavlja bias čvora, ali pošto je on nezavisan od ulaznih vrednosti neuronske mreže možemo ga ignorisati u daljim računanjima.

Na sličan način prethodnom možemo dobiti formulu za izlazni čvor y_1 . Pošto je y_1 u poslednjem sloju on ujedno predstavlja verovatnoću pripadnosti ulazne slike nekoj od klasa. Vrednost čvora y_1 možemo predstaviti sledećom formulom:

$$y_1 = a_4 w_{84} + a_5 w_{85} + a_6 w_{86} + a_7 w_{87}$$

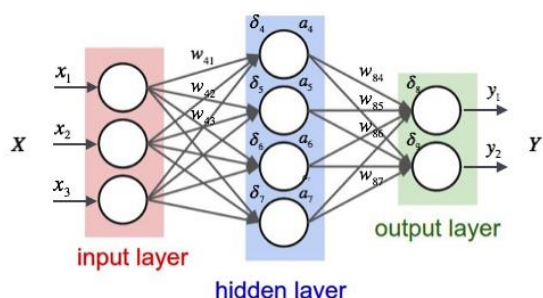
Uvrštavanjem formula za čvorove a_n možemo dobiti vrednost čvora y_1 u obliku zbira ulaznih vrednosti iz početnog sloja pomnoženih sa odgovarajućim koeficijentima:

$$y_1 = K_1 x_1 + K_2 x_2 + K_3 x_3$$

gde je K_n zbir proizvoda težina na svkom putu iz x_n u y_1 . Na primer:

$$K_1 = w_{41} w_{84} + w_{51} w_{85} + w_{61} w_{86} + w_{71} w_{87}$$

Na osnovu ovih formula se može izračunati uticaj svake ulazne vrednosti x_n pomoću njegovog koeficijenta K_n . To znači da na osnovu vrednosti koeficijenta možemo odrediti vrednosti ulazne slike tako da maksimizujemo izlaz u čvoru y_1 . Jednostavnije rečeno, ako je koeficijent negativan broj onda za ulaz treba uzeti što manji broj (u domenu vrednosti piksela slike to je 0) a ako je pozitivan onda što veći (u domenu piksela to je 255).



Slika 1 - Primer jednostavne neuronske mreže

IMPLEMENTACIJA REŠENJA

Programska implementacija za do sada navedene formule je veoma komplikovana i neefikasna, ponaročito u slučaju većih neuronskih mreža. Takođe se nailazi na problem *underflow*-a pošto su koeficijenti često brojevi blizu nuli, i njihovim neprestanim množenjem se dobijaju brojevi koji su previše mali da bi bili ispravno reprezentovani u računaru.

Mnogo boljim rešenjem se predstavio algoritam u kojem se jedan po jedan menjaju vrednosti piksela u slici, a zatim posmatra izmena u izlaznoj vrednosti neuronske mreže. Time je moguće utvrditi uticaj svakog pojedinačnog piksela na rezultat, pa se na osnovu toga formira potpuna slika. Mana ove implementacije jeste što ona ne funkcioniše podjednako dobro na neuronskim mrežama koje implementiraju max-pooling mehanizme, ali se ovaj algoritam i dalje pokazao kvalitetnijim. Zbog toga se on i koristi za implementaciju rešenja.

REZULTATI

Tokom izvršavanja algoritma postalo je veoma je očigleno da slike koje se generišu ne izgledaju kao klase koju bi trebale predstavljati, iako su dobijene slike davale bolje rezultate čak i od slika iz datasetova koji se koriste za treniranje. To delimično i ukazuje na to da su korišćene neuronske mreže nedovoljno kompleksne.

Za testiranje algoritma smo koristili tri različite neuronske mreže:

- Za prepoznavanje lica^[1]. Prilikom pokretanja algoritma dobije se slika za koju mreža, sa 100% sigurnosti, ustanovi da je lice.



Slika 2 - Izgenerisana slika za klasu lice

- Neuronska mreža za prepoznavanje cifara^[2], isto kao i sa prethodnim primerom za svaku cifru algoritam generiše sliku za koju neuronska vrati 100%-nu sigurnost.



Slike 3 i 4 - Izgenerisane slike za cifre 0 i 8

- Konvolutivna neuronska mreža sa max-pooling za prepoznavanje cifara^[2]. Rezultati za ovu mrežu su bili uočljivo lošiji od predhodnih samo je za jednu klasu uspešno generisana slika koja pripada toj klasi sa sigurnošću iznad 90%, za ostatak klasa su vrednosti uglavnom ispod 50%.

Kao što je i očekivano, dobijene slike ne pokazuju podjednako dobre rezultate kada se primene na neke druge klasifikatore, čak i ako je su to neuronske mreže različite strukture, trenirane nad istim datasetom. Ovo je najverovatnije rezultat uske povezanosti generisane slike sa unutrašnjom strukturom neuronske mreže.

REFERENCE:

- [1] - Dataset korišćen za treniranje: Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. - **Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments.**
- [2] - Dataset korišćen za treniranje: **The MNIST database of handwritten digits**