airs(test_sample2, token_sed

Prikaz slike na osnovu pročitanog teksta

Autor: Bojana Kliska RA98/2016

Asistent: Stefan Anđelić

Motivacija

Detekcija i identifikacija karaktera jednog jezika sa vizuelnog ulaza predstavlja jednu od aktuelnih tema kada se govori o vizuelnom prepoznavanju u oblasti veštačke inteligencije. Pored toga, takođe značajna tema jeste i klasifikacija reči jednog jezika na osnovu semantičkih karakteristika, na vrste reči. Na osnovu prepoznate i odabrane određene vrste reči, generisanje slike bilo bi značajno, na primer za generisanje vizuelnog rečnika ili kao olakšica osobama koje imaju poteškoće prilikom čitanja teksta ili razumevanja pročitanog.

Skupovi podataka

Za realizaciju celokupne ideje bilo je potrebno napraviti dva modela:

- 1) Model koji prepoznaje karaktere sa slike teksta,
- 2) Model koji prepoznaje vrstu reči.

Stoga je bilo potrebno pripremiti i dva skupa podataka za obučavanje i testiranje modela.

Za prvi model za obučavajući skup koristila se slika štampanih slova engleskog alfabeta, malih i velikih, ali samo jednog fonta.

Samim tim, sve slike iz test skupa bile su napisane istim fontom.

ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz

Za drugi model za POS tagging (Part-of-speech tagging) kao obučavajući skup koristio se Penn Treebank tagset, jedan od najpopularnijih tag setova za engleski jezik.

Model je bio testiran i na delu rečenica preuzetih iz pomenutog tag seta, ali i na proizvoljnim rečenicama, kao i na rečenicama dobijenim nakon izvršenog prepoznavanja karaktera sa slike.

Obučavanje modela

Za obučavanje modela koji prepoznaje karaktere korišćena je veštačka neuronska mreža sa jednim ulaznim i jednim izlaznim slojem.

Za obučavanje modela za POST korišćena je LSTM sa 4 sloja – 3 skrivena i jednim izlaznim slojem.

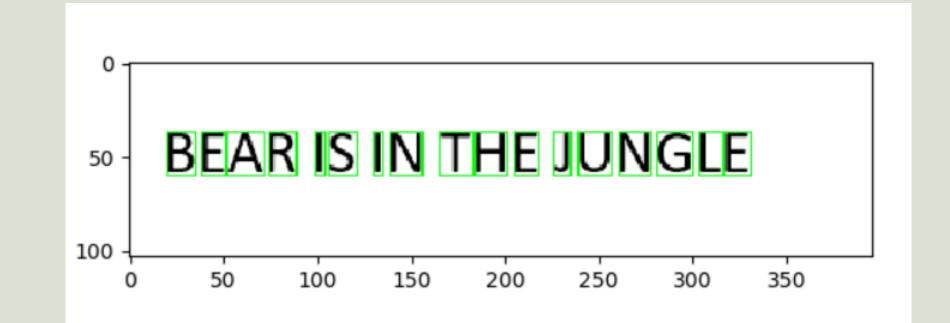
Alphabetical list of part-of-speech tags used in the Penn Treebank Project:

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential there
5.	FW	Foreign word
б.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.		Proper noun, plural
16.		Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
		Possessive pronoun
20.		Adverb
21.	RBR	Adverb, comparative
		Adverb, superlative
		Particle
24.	SYM	Symbol
25.	TO	to
26.	UH	Interjection
		Verb, base form
28.		Verb, past tense
29.		Verb, gerund or present participle
30.		Verb, past participle
31.		Verb, non-3rd person singular present
32.		Verb, 3rd person singular present
33.		Wh-determiner
34.	WP	Wh-pronoun
		Possessive wh-pronoun
36.		Wh-adverb

Pretprocesiranje podataka

Pretprocesiranje podataka prvog modela za detekciju karaktera podrazumevalo je:

- Učitavanje slike,
- Pretvaranje slike u boji u greyscale,
- Korišćenje threshold-a,
- Pronalaženje kontura od interesa (karaktera na slici),
- Resize-ovanje slika od kontura ,
- Pretvaranje matrica slika u vektore kao pogodne ulazne podatke za neuronsku mrežu.



Pretprocesiranje podataka drugog modela za POS tagging podrazumevalo je:

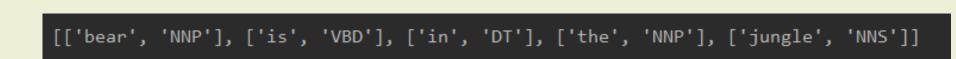
- Preuzimanje označenih rečenica iz Pen Treebank tag set-a,
- Podelu rečenica na train i test skup podataka,
- Odvajanje reči od tagova,
- Formiranje setova jedinstvenih reči i tagova,
- Formiranje dictionary-ja gde se na svaku od jedinstvenih reči/tagova mapira broj, te se tako reči i tagovi predstavljaju u vidu matrica brojeva, pogodnih za ulaz u neuronsku mrežu,
- Za reči iz seta koje nemaju tag dodat je tag "-OOV-" (out of vocabulary),
- Takođe je dodat i tag "-PAD-" on služi kako bi se popunila prezna mesta u kratkim rečenicama (nisu sve rečenice istih dimenzija, padding omogućava da budu, tako što na osnovu dužine najduže rečenice dodaje sa desna -PAD- tag, tj. broj 0 svim rečenicama koje su od nje kraće).

```
BEAR is iN ThE juNGLE
bear is in the jungle
['bear', 'is', 'in', 'the', 'jungle']
[[[3.5619028e-03 1.8585527e-02 8.8323550e-03 ... 4.0594011e-04
3.5091895e-03 5.6388666e-04]
[1.2965583e-02 4.0914700e-03 7.0353257e-03 ... 2.0638413e-03
1.0381277e-02 1.3163959e-03]
[1.2589888e-05 2.4090735e-07 2.6385826e-05 ... 2.4263536e-04
1.7300947e-05 1.4411454e-06]
```

Testiranje modela

Nakon dobijenih rezultata može se zaključiti da:

- 1) Model za prepoznavanje karaktera dobro radi kada mu je ulazna slika napisana velikim slovima.
- a) Međutim ima problema prilikom prepoznavanja malih slova. Ovaj problem uzrokuju različite veličine slova i slika na kojima se testira, jer zbog različitih veličina ne može adekvatno da pronađe konture od interesa
- b) Takođe, pošto je model obučavan samo na jednom fontu, za različite fontove ne daje dobre rezultate.
- 2) Model za POS tagging daje veoma dobre rezultate, sa tačnošću i do 99 procenata.
- a) Međutim, jedna od posledica velike tačnosti modela jeste to što on prilikom obučavanja uzima u obzir i padding deo rečenice, koji u velikom broju slučajeva čini većinu ulaznog podatka. Padding deo je lako prepoznati i naučiti.
- b) Pored toga primećene su situacije u koji takođe pravi grešku. Naime, ukoliko je u pitanju rečenica: "Running is very important for me.", prvoj reči u rečenici: "running" trebalo bi da dodeli teg "NN" što predstavlja imenicu (Noun, singular or mass). Međutim, on toj reči dodeli tag "VBG" (Verb, gerund/present), što morfološki može biti tag te reči, ali u toj konkretnoj rečenici, semantički, nije.
- c) Takođe, nekada se desi da određene znake interpunkcije ne prepozna dobro, na primer, često znak "; " prepozna kao ": ".



Prikaz slike na osnovu odabranih reči

Nakon obrade slike i dobijenog teksta, taj tekst se šalje modelu za određivanje vrste reči. Nakon toga, svakoj reči iz prosleđenog teksta, dodeljuje se tag koji opisuje vrstu reči kojoj je prediktovano da reč pripada.

Nakon toga, potrebno je da se uradi Google image download za one reči koje su dobile odabrane tagove. Pošto je ideja bila prikazati samo imenice, odabrani tagovi su biti: "NN", "NNP", "NNS" i "NNPS". Kada se odradi download slika sa Google-a za željene reči, formiraju se folderi unutar root foldera projekta sa nazivom istim kao prosleđena reč, u kojem se čuvaju skinute slike.

