

Generiranje imena naselja pomocu LSTM mreže

Antonio Čogelja Morena Granić Fran Lubina Iva Jurković Jakov Juvančić Matej Logarušić

Sažetak—Cilj projekta je LSTM rekurzivna neuronska mreža na razini znakova koja generira realistična imena hrvatskih naselja. Fokus projekta je treniranje i razvijanje neuronske mreže za generiranje realističnih imena hrvatskih naselja. Korištenjem LSTM mreže, koja je prilagođena za analizu sekvencijskih podataka, cilj je razviti model sposoban za učenje jezičnih obrazaca i struktura iz postojećih imena naselja. Svrha mreže je generiranje novih imena temeljenih na tim naučenim obrascima, pri čemu se zadržavaju jezične i strukturne zakonitosti specifične za taj kontekst. Željena točnost modela η je $\lim_{\tau \rightarrow 0} \eta = 0.5$

Index Terms—Naselje, LSTM, rekurzivne mreže, neuronske mreže-

I. UVOD

Ishod projekta je LSTM rekurzivna neuronska mreža na razini znakova koja generira realistična imena hrvatskih naselja. Mreža radi sa vektorima koji predstavljaju slova hrvatske abecede proširene specijalnim znakovima $\Sigma = \{\text{hrv. abeceda}\} \cup \{\text{start}, " \setminus 0" \}$.

Ulaz mreže je one-hot vektor $\mathbf{x}^{(t)}$ dimezije $|\Sigma| = 30 + 2$.

$$\mathbf{x}_i^{(t)} = \begin{cases} 1, & \text{ako } i = j \\ 0, & \text{inače} \end{cases} \quad (1)$$

Izlaz dobiven na kraju pojedinog vremenskog koraka t je vektor vjerojatnosti pojave pojedinog znaka abecende.

$$\hat{\mathbf{y}}^{(t)} = \begin{bmatrix} p(c_0) \\ p(c_1|c_0) \\ \vdots \\ p(c_{|\Sigma|-1} | \bigcap_{i=0}^{|\Sigma|-2} c_i) \end{bmatrix} \quad \text{Gdje } c \in \Sigma \quad (2)$$

Vjerojatnosti su dobivene softmax funkcijom parametriziranom hiperparametrom temperature τ .

Na temelju tih vjerojatnosti se uzorkuje konačni izlazni vektor $\mathbf{y}^{(t)}$, odnosno t -ti znak u imenu naselja.

$$\mathbf{y}^{(t)} \sim \hat{\mathbf{y}}^{(t)} = \sigma_{\tau}(f(\mathbf{x}^{(t)}; \boldsymbol{\theta})) \quad (3)$$

$f(\mathbf{x}; \boldsymbol{\theta})$ predstavlja ukupno djelovanje ćelija modela nad njenim ulazom parametrizirano hiperparametrima modela $\boldsymbol{\theta} = [|\mathbf{a}| \quad \mu \quad \tau]$

(opisani u poglavlju III-B) Temperaturno uzorkovanje je izabrano, jer omogućava eksperimentiranje i generiranje zanimljivih toponima. Izlaz mreže je niz znakova $\{\mathbf{y}^{(t)}\}_{t=0}^{T-1}$, odnosno ime naselja. Željena točnost modela η je $\lim_{\tau \rightarrow 0} \eta = 0.5$

II. PREGLED LITERATURE

III. OPIS IMPLEMENTIRANE LSTM MREŽE

Fokus projekta je treniranje i razvijanje neuronske mreže za generiranje realističnih imena hrvatskih naselja. Korištenjem LSTM mreže, koja je prilagođena za analizu sekvencijskih podataka, cilj je razviti model sposoban za učenje jezičnih obrazaca i struktura iz postojećih imena naselja. Svrha mreže je generiranje novih imena temeljenih na tim naučenim obrascima, pri čemu se zadržavaju jezične i strukturne zakonitosti specifične za taj kontekst. LSTM ćelija i mreža je implementirana u radnom okviru pyTorch.

Dizajn mreže i podešavanje hiperparametara se odvija paralelno sa implementacijom mreže u radnom okviru Keras.

A. Arhitektura

–

B. Hiperparametri

- 1) Dimenzija skrivenog stanja: $|\mathbf{a}|$
- 2) Stopa učenja: μ
- 3) Temperatura: τ
- 4) Broj LSTM ćelija

C. Ćelija

Ćelija izgleda ovako.

D. Treniranje

BPTT je korišten kao algoritam učenja. Kao funkcija gubitka koristi se unakrsna entropija.

$$L_i = - \sum_{t=0}^{T-1} \mathbf{x}_i^{(t)} \cdot \log(\hat{\mathbf{y}}_i^{(t)}) \quad (4)$$

BPTT je u biti propagirani stohastički gradijentni spust. Težine se mijenjaju prema pravilu:

$$\Delta w = -\eta \frac{\partial L}{\partial w} \quad (5)$$

IV. OPIS EKSPERIMENTALNIH REZULTATA

a) optimiranje hiperparametara: –

Tablica I
TABLE TYPE STYLES

Table Head	Table Column Head		
	Table column subhead	Subhead	Subhead
copy	More table copy ^a		

^aSample of a Table footnote.

A. Usporedba rezultata

–

LITERATURA

- [1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.