

# Language translation model using neural machine translation with seq2seq architecture

---

JELENA ZARIC 79/19

# Uvod

---

- Neuronsko mašinsko prevodenje (NMT) pristup učenju od početka do kraja za automatizovani prevod. Njegova moć dolazi iz direktnog učenja mapiranjem ulaznog teksta u povezani izlazni tekst.
- Dokazano efikasniji od tradicionalnog prevodenja zasnovanog na frazama, koje zahteva mnogo više truda u dizajniranju modela. S druge strane, NMT modeli su skupi za obuku, posebno na velikim skupovima podataka za prevod. Takođe značajno sporiji zbog velikog broja parametara.
- Drugi ograničavajući faktori su poteškoće u prevodjenju svih delova ulazne rečenice. Postoje neka rešenja da bi se prevazišli ovi problemi, npr. korišćenje mehanizma pažnje za kopiranje retkih reči.

# Opis problema i skup podataka

---

- Francusko-engleski rečnik - skup podataka koji sadrži oko 170 000 linija, u svakoj liniji se nalazi najpre rečenica na engleskom, a zatim njen prevod na francuski jezik.
- Za potrebe projekta uzet je manji podskup od 20 000 linija, radi bržeg treninga.

Go.	Va !
Hi.	Salut !
Run!	Cours !
Run!	Courez !
Who?	Qui ?
Wow!	Ça alors !
Fire!	Au feu !
Help!	À l'aide !
Jump.	Saute.
Stop!	Ça suffit !
Stop!	Stop !
Stop!	Arrête-toi !
Wait!	Attends !

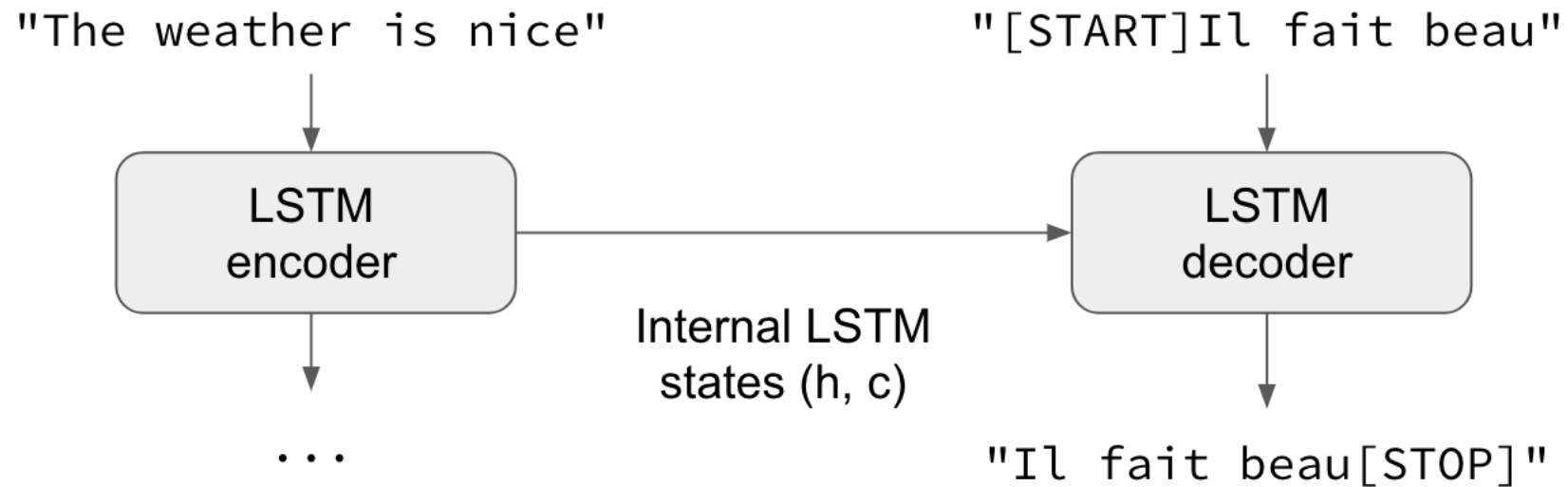
# Model mreže

---

- Ovaj model mreže je implementacija sekvencijskog prevodilačkog sistema koji koristi rekurentne neuronske mreže (LSTM) za prevodenje sa jednog jezika na drugi.
- Model se sastoji od enkodera i dekodera. Enkoder prima ulaznu sekvencu i obrađuje je putem LSTM sloja da bi izvukao relevantne zaključke. Nakon obrade, deo enkodera se odbacuje, a zadržavaju se unutarnja stanja. Dekoder koristi unutarnja stanja kao početna stanja i generiše izlaznu sekvencu na ciljanom jeziku. Dropout za regularizaciju.
- Na kraju, model koristi softmax aktivacionu funkciju nad poslednjim slojem kako bi generisao konačan prevedeni tekst.

# Model mreže

---



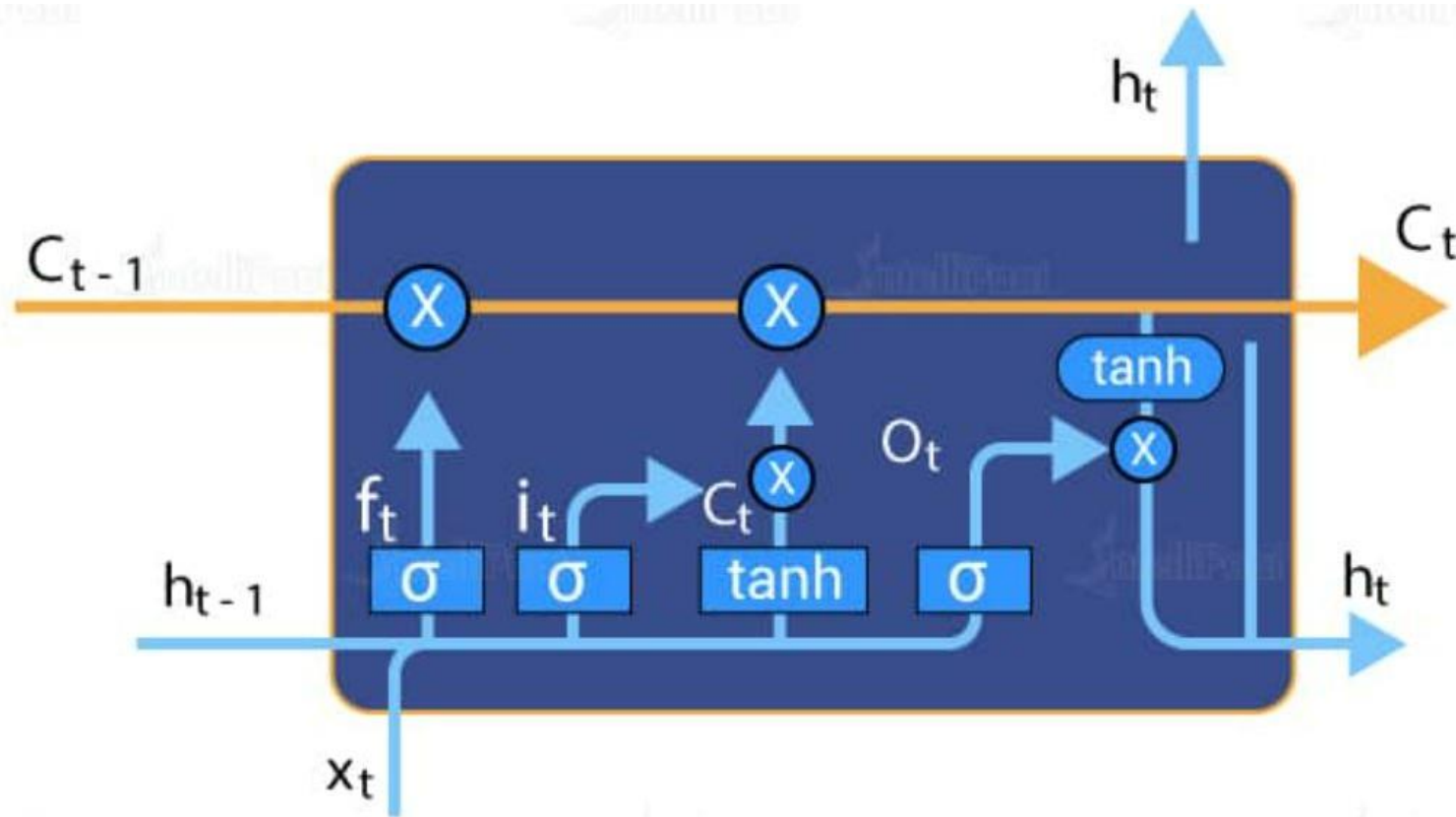
# Šta je LSTM?

---

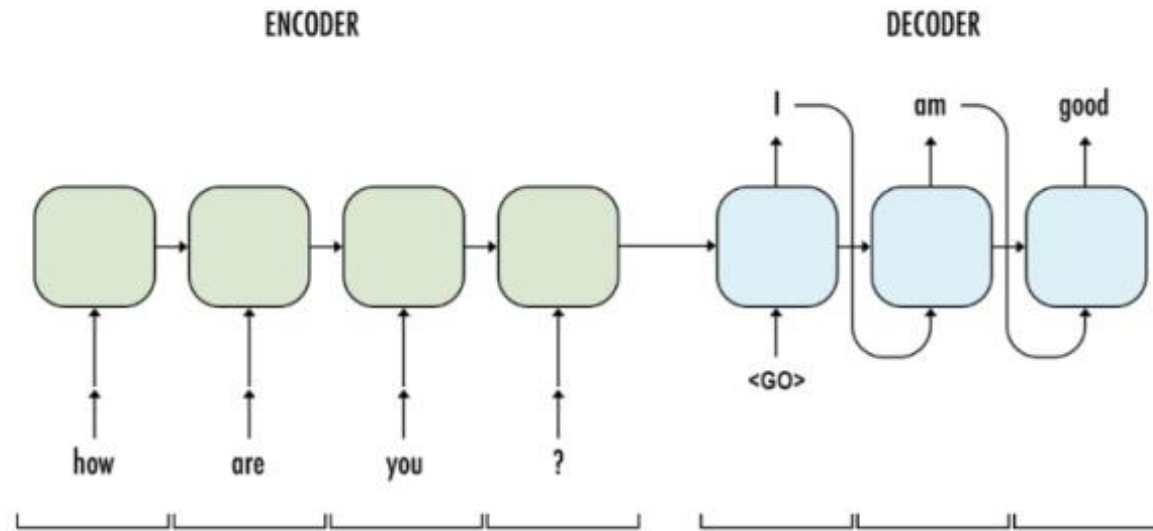
- LSTM (long short-term memory networks) je skraćenica za mreže kratkoročnog pamćenja, koje se koriste u oblasti dubokog učenja. To je niz rekurentnih neuronskih mreža (RNN) koje mogu da uče zavisnosti dugoročno, posebno u problemima predviđanja sekvenci.
- Sa povratnim vezama, imaju sposobnost da obrade čitave nizove podataka. Posebna je vrsta RNN koja pokazuje odlične performanse u raznim oblastima (prepoznavanju govora, mašinskom prevođenju... ).

# Šta je LSTM?

- Logika koja stoji iza LSTM



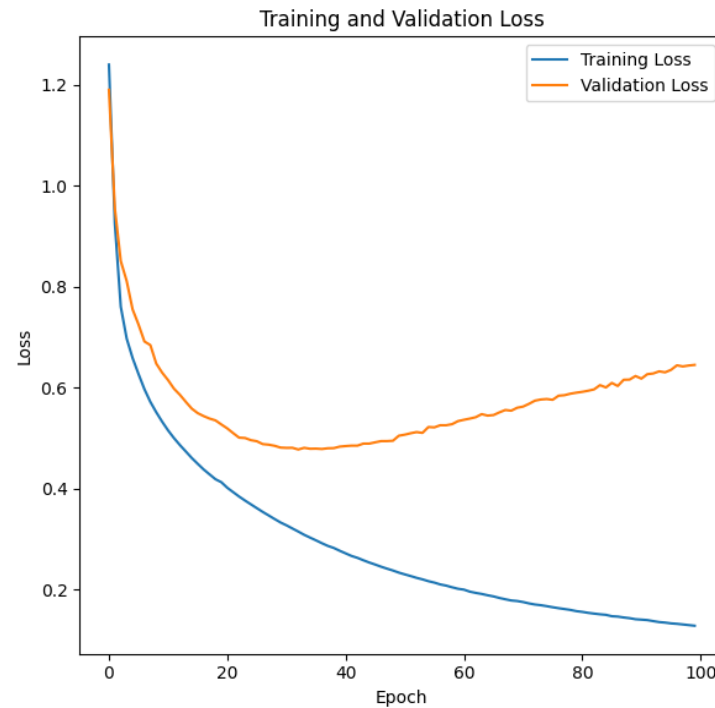
```
encoder_input_data = np.zeros((len(input_texts), max_encoder_seq_length, num_encoder_tokens),dtype='float32')
decoder_input_data = np.zeros((len(input_texts), max_decoder_seq_length, num_decoder_tokens),dtype='float32')
decoder_target_data = np.zeros((len(input_texts), max_decoder_seq_length, num_decoder_tokens),dtype='float32')
```





# Testiranje i rezultati

- optimizator “rmsprop”
- categorical crossentropy
- batch size = 64
- 100 epoha
- 20% podataka za validaciju  
(validation split = 0.2)



# Moguća poboljšanja

---

- Osnovna seq2seq arhitektura nije sposobna za hvatanje konteksta i jednostavno uči kako da mapira pojedinačne ulaze na pojedinačne izlaze.
- Word embedding (ugradnja reči) – Bolje je koristiti jer duboki modeli za učenje rade s brojevima. Zbog toga pretvaramo naše reči u odgovarajuće numeričke vektorske reprezentacije, a ne samo u niz brojeva. Encoder barata vektorima, ne pojedinačnim vrednostima.
- Sloj za ugradnju reči mapira sekvencu indeksa reči u vektore za ugradnju i uči ugrađivanje reči tokom treninga. Ovakvom sloju treba određen alat za duboko učenje. Na ovaj način model može naučiti određene veze između reči.

Hvala na pažnji!