

DETEKCIJA SARKAZMA U NASLOVIMA NOVINSKIH ČLANAKA UPOTREBOM NEURONSKIH MREŽA**NEWS HEADLINES SARCASM DETECTION USING NEURAL NETWORKS**Aleksandar Vujinović, *Fakultet tehničkih nauka, Novi Sad***Oblast – RAČUNARSTVO I AUTOMATIKA**

Kratak sadržaj – Sarkazam je način komunikacije koji predstavlja suprotnost od pravog značenja sa ciljem da se na duhovit način nešto kritikuje. Vrlo je bitno da sarkazam u komunikaciji bude shvaćen na pravi način kako ne bi došlo do nerazumevanja. U pisanoj reprezentaciji je sarkazam teže uočljiv jer nedostaju svi neverbalni signali. U ovom radu prikazana je detekcija sarkazma u novinskim naslovima korišćenjem: (1) modela zasnovanih na neuronskim mrežama koji koriste vektorske reprezentacije teksta i (2) modela zasnovanih na transformer arhitekturi. Kao najbolji od modela zasnovanih na neuronskim mrežama pokazao se model sa LSTM (Long Short-Term Memory) i konvolutivnim slojevima, postigavši tačnost od 86%. Transformer modeli su nadmašili rezultate grupe modela zasnovanih na neuronskim mrežama i vektorskoj reprezentaciji teksta. Među njima, najbolje rezultate pokazao je model roBERTa sa tačnošću od 94%.

Ključne reči: detekcija sarkazma; novinski članci; duboko učenje; neuronske mreže.

Abstract – Sarcasm is the use of a remark that means the opposite of the real meaning, made to criticize something in a humorous way. It's essential to understand sarcasm to avoid misunderstanding. Detecting sarcasm in the written text is problematic since all non-verbal signals are missing. In this paper, I present sarcasm detection in news headlines using: (1) embedding-based neural networks and (2) transformer-based models. LSTM (Long Short-Term Memory) with convolutional layers achieved the best performance among embedding-based models, reaching 86% accuracy, whereas roBERTa reached 94% accuracy within transformer-based models. All the transformer models outperform embedding-based models.

Keywords: sarcasm detection; news articles; deep learning; neural networks.

1. UVOD

Detekcija sarkazma u tekstu je jedan od problema u NLP (Natural Language Processing) oblasti. Sarkazam predstavlja specifičan slučaj analize sentimenta, odnosno, klasifikacionog problema. Detekcija sarkazma se ne mora posmatrati kao zaseban alat, već kao dopuna postojećim algoritmima, kako bi se unapredile njihove performanse.

NAPOMENA:

Mentor ovog rada, koji je sažetak master rada, bila je prof. Jelena Slivka.

U ovom radu je predstavljeno više modela dubokog učenja za detekciju sarkazma u naslovima novinskih članaka. Na osnovu datog ulaznog teksta sistem vraća odgovor da li je tekst sarkastičan ili nije. U osnovi svih modela se nalaze neuronske mreže. Korišćene modele možemo podeliti u dve klase na osnovu arhitekture:

1. modeli zasnovani na neuronskim mrežama koje koriste vektorske reprezentacije reči (eng. *word embedding*) i
2. modeli zasnovani na transformer arhitekturi.

U okviru prve grupe modela, isprobane su potpuno povezane neuronske mreže i RNN (Recurrent Neural Network) mreže koje koriste LSTM (Long Short-Term Memory) ili GRU (Gated Recurrent Unit) slojeve. U okviru druge grupe modela, isprobani su transformer modeli koji su pokazali bolje performanse pri detekciji sarkazma: BERT [1], roBERTa [2], GPT-2 [3] i T5 [4] pretrenirani transformeri.

Svi modeli druge grupe nadmašuju performanse modela iz prve grupe. U prvoj grupi modela, najbolje performanse je postigao model sa LSTM i konvolutivnim slojevima od 86% tačnosti (eng. *accuracy*). U drugoj grupi modela, najbolje performanse je postigao roBERTa model od 94% tačnosti. To je veća tačnost za 8% u odnosu na najbolji model zasnovan na vektorskoj reprezentaciji reči.

2. SRODNA ISTRAŽIVANJA

Kako je fokus ovog rada na detekciji tekstualne reprezentacije sarkazma u naslovima novinskih članaka, u ovom poglavlju su predstavljeni najbitniji radovi sa istim istraživačkim fokusom.

Arhitektura predložena u radu [5] koristi arhitekturu koja kao ulaz prima naslove, a kao izlaz vraća binarnu vrednost (1 ukoliko je naslov sarkastičan, a 0 ukoliko nije). U model je dodat bidirekcioni LSTM modul kako bi enkodirao levi (i desni) kontekst u rečenici. Ovaj LSTM modul je dopunjen *attention* modulom kako bi se re-evaluirale težine enkodiranog konteksta. Autori rada su trenirali model na balansiranom skupu podataka, gde je, nakon podešavanja hiper-parametara modela, dobijen rezultat od 89,7% tačnosti na test skupom. Ovaj rezultat predstavlja poboljšanje u odnosu na polazni model konvolutivne neuronske mreže korišćen u ovom istraživanju, gde se pomoću *embedding* sloja dobila tačnost 84,88% nad test skupom.

U radu [6] je prikazana detekcija sentimenta, emotikona i sarkazma korišćenjem neuronske mreže trenirane nad Twitter [7] postovima. Predložena arhitektura se sastoji od *embedding* sloja, dva bidirekcionala LSTM sloja i *attention* sloja. U radu je ilustrovano da ovakva arhitektura jeste

pogodna za *twitter* postove različitih dužina, koji se sastoje od jedne ili više rečenica. Nad skupom podataka korišćenim u radu [6], postignuta je prosečna F1 mera od 69%. Pokazali su da LSTM i *attention* sloj proizvode boljim performansama u odnosu na polazne modele koji nisu koristili navedene slojeve.

Rad [8] prikazuje poređenje različitih modela nad više skupova podataka koji obuhvataju objave na *Twitter*-u, *Reddit*-u [9] i naslove novinskih članaka. Autori su poredili neuronske mreže različitih arhitektura. Utvrdili su da mreža koja sadrži LSTM sloj postiže bolje performanse u odnosu na mrežu koja koristi isključivo konvolutivne slojeve. Takođe su utvrdili da kombinacija LSTM i konvolutivnih slojeva daje još bolje rezultate. Kada su na LSTM sloj nadovezali i *attention* sloj, dobili su još bolje rezultate. Najbolje rezultate dobili su korišćenjem transformer modela. Prikazano je da modeli zasnovani na vektorskoj reprezentaciji reči ostvaruju lošije rezultate od transformer modela, konkretno, BERT modela. Pokazali su i da će transformer model sa više glava da postiže bolje rezultate, jer može više parametara da nauči. Nad istim skupom podataka, naslovima novinskih članaka, su postigli tačnost i F1 meru od 92%.

3. SKUP PODATAKA

Skup podataka korišćen u ovom radu [5] se sastoji se od sarkastičnih i nesarkastičnih naslova novinskih članaka.

U skupu podataka ne postoje polja koja nisu popunjena. Broj redova (novinskih članaka) koji se nalazi u skupu podataka iznosi 26.709, a broj jedinstvenih reči je 263.016. Broj sarkastičnih i nesarkastičnih članaka je približno izbalansiran: postoji 11.724 sarkastična i 14.985 nesarkastičnih članaka.

Da bi se naslovi novinskih članaka pretvorili u vektorsku reprezentaciju, neophodno je pretprocesirati podatke. Isprobane metode čišćenja podataka su:

- Prebacivanje čitavog teksta u mala slova.
- Izbacivanje brojeva iz teksta.
- Transformacija apostrof reči - Na primer, „isn't” je transformisano u „is not”, „couldn't” je transformisano u „could not”.
- Uklanjanje interpunkcije (interpunkcije).
- Uklanjanje *stop* reči.
- Primena steminga (eng. *stemming*).
- Primena lematizacije (eng. *lemmatization*).

Uklanjanje *stop* reči, uklanjanje brojeva, uklanjanje znakova interpunkcije, steming i lematizacija se nisu pokazale kao uspešne metode. Svaka od ovih metoda je u nekoj meri doprinela da performanse modela opadnu. Nasuprot ovim metodama, metode kao što su svođenje reči na mala slova, uniformni apostrof, i transformacija apostrof reči su se pokazale kao metode koje su u kombinaciji doprinele da se poboljšaju performanse modela.

Drugi korak u pretprocesiranju je tokenizacija. Za BERT model postoji već pretrenirani *tokenizer* koji svaku reč pretvara u broj, koristeći pretrenirani BERT rečnik, i dodaju se specijalni karakteri koji se isto tokenizuju, (predstavljaju brojevima). Na taj način se za svaki naslov napravi vektorska reprezentacija, odnosno tenzor. Ovako pretprocesiran skup podataka je ulaz u BERT model.

roBERTa model koristi iste korake kao i BERT model sa jedinom razlikom da nije korišćen pretrenirani BERT *tokenizer* već pretrenirani roBERTa *tokenizer*. Analogno važi za T5 model i analogno važi za modele zasnovane na vektorskoj reprezentaciji reči. Za GPT-2 bilo je neophodno ručno dodati specijalne tokene PAD i CLS. Pošto se CLS nalazi na kraju sekvence, a naš zadatak je klasifikacioni problem, neophodno je promeniti stranu za peding i odsecanje, da bude sa leve strane, a ne sa desne.

4. METODOLOGIJA

Zadatak ovog rada je da se klasifikuju naslovi novinskih članaka u sarkastične i nesarkastične naslovi. U tu svrhu korišćeni su modeli zasnovani na neuronskim mrežama. Tekstualna reprezentacija nije pogodna kao ulaz za neuronske mreže. Zbog toga smo svaki naslov pretvorili u vektorsku reprezentaciju. Isprobani su modeli zasnovani na neuronskim mrežama koji koriste vektorske reprezentacije reči i modeli zasnovani na transformer arhitekturi.

4.1. Modeli zasnovani na vektorskoj reprezentaciji reči

Kao osnova za poređenje sa svim drugim modelima, napravljena je jednostavna potpuno povezana neuronska mreža. Izlazni sloj sadrži jedan neuron sa sigmoid aktivacionom funkcijom, jer imamo dve moguće klase: sarkastično ili nesarkastično. Svi slojevi u ovoj mreži su se obučavali. Osnovni model se sastoji od sledećih slojeva:

- *Embedding* sloj ulazne dimenzije 10000 i izlazne dimenzije 16,
- *Pooling* sloj, odnosno, *GlobalAveragePooling*,
- *Dense* sloj, jedan skriveni potpuno povezani sloj sa 24 neurona.

Zatim smo osnovni model proširili dodatnim slojevima, i naštimovali hiper-parametre modela: dimenzija ulaza, odgovara broju reči u rečniku, i dimenzija izlaza u *Embedding* sloju, korak u optimizacionoj funkciji (eng. *learning rate*) i broj neurona u svakom sloju.

Takođe je rekonstruisan model po ugledu na rad [5]. Bitna razlika u ovom modelu u odnosu na prethodna dva je što ovaj model jeste RNN model koji koristi i konvolutivne slojeve. Iskoristili smo konvolutivne slojeve i bidirekcionu LSTM sloj koji je pogodan za rad sa sekvencijalnim podacima, pa i sa tekstom koji je sekvenca reči.

Zatim su isprobana tri različita RNN modela. Sva tri modela implementirana su koristeći *Keras* [10] biblioteku. Hiper-parametri modela su optimizovani koristeći *Keras Tuner* [11] biblioteku.

Prvi RNN bazirani model koristi *embedding* sloj koji nije pretreniran, a težine u *embedding* sloju će se odrediti u procesu obučavanja modela. Zatim sledi *pooling* sloj, pa tri bidirekciona LSTM sloja, gde se pre svakog LSTM sloja nalazi *dropout* sloj. Na ove slojeve nadovezuje se izlazni klasifikacioni sloj.

Drugi RNN model se od prvog razlikuje što umesto tri LSTM sloja koristi tri GRU (*Gated Recurrent Unit*) sloja.

Treći model se dosta razlikuje od prva dva modela. Umesto nepretreniranog *embedding* sloja, učitani su pretrenirani *embedding* sloj izlazne dimenzije 200, treniran nad *Twitter* skupom podataka. Za ostale slojeve isprobane su različite kombinacije i na kraju je najbolje rezultate dala mreža koja

se sastoji od kombinacije konvolutivnih, potpuno povezanih, *pooling*, *dropout* i više bidirekcionih GRU slojeva. Prvo je isprobano da se *embedding* sloj zamrzne, odnosno da se težine ne menjaju u procesu obučavanja.

Na taj način vektorska reprezentacija reči ostaje onakva kakva je prethodno naučena. Ovakva mreža davala je dosta lošije rezultate u odnosu na mnogo jednostavnije modele, pa je zbog toga odlučeno da se *embedding* sloj ipak dotrenira u procesu obučavanja.

4.2. Modeli zasnovani na transformer arhitekturi

U radu su isprobani: (1) enkoderski modeli: BERT [1], roBERTa [2], (2) dekoderski modeli: GPT-2 [3] i (3) model koji ima enkoder-dekoder arhitekturu T5 [4].

4.2.1. BERT

BERT model jezika je zasnovan na transformer arhitekturi [12]. Za naše potrebe smo koristili pretrenirani *bert-base-uncased* model iz *Huggingface* [13] biblioteke. Na poslednji skriveni sloj je dodat izlazni klasifikacioni sloj sa 2 neurona, kako bi se mogla izvršiti klasifikacija na sarkastične i nesarkastične naslove novinskih članaka. Model kao ulaz prima naslove novinskih članaka u vektorskoj reprezentaciji. Prvo sam primenio pretprocesiranje, a zatim trenirao BERT modela. Hiper-parametre modela sam naštimovao po ugledu na rad [12].

4.2.2. roBERTa

roBERTa model se zasniva na BERT modelu. Koristio sam pretrenirani model iz *Huggingface* biblioteke. Da bi se naslovi pretvorili u vektorsku reprezentaciju, neophodno je pretprocesirati podatke. Nakon završenog pretprocesiranja, sledi trening roBERTa modela. Hiper-parametri modela su naštimovani po ugledu na rad [14].

4.2.3. T5

T5 model se za razliku od BERT i roBERTa modela sastoji od enkoderskog i dekoderskog dela. Pošto naš zadatak spada u klasifikacioni problem, dekoderski deo je višak. Korišćen je pretrenirani enkoderski deo modela na koji je nadovezan jedan izlazni klasifikacioni sloj sa 2 neurona.

4.2.4. GPT-2

GPT-2 transformer model se za razliku od BERT, roBERTa i T5 modela sastoji samo od dekoderskog dela. Model je treniran za zadatak predikcije sledeće reči, preciznije sledećeg tokena u sekvenci.

Zbog prirode zadatka za koji je model treniran, model može da pristupi samo rečima sa leve strane u odnosu na trenutnu reč, za razliku od bidirekcionih modela koji mogu da pristupaju rečima sa obe strane. Kontekst odnosno podaci o celoj sekvenci se nalaze u CLS tokenu, a CLS token jeste poslednji token u sekvenci, za razliku od enkoderskih modela.

5. REZULTATI I DISKUSIJA

Skup podataka je podeljen na trening, validacioni i test skup u odnosu 60/20/20. Kao mera evaluacije korišćeni su tačnost i F1 mera. Prikaz rezultata biće podeljen u dve grupe: modeli zasnovani na neuronskim mrežama koji koriste vektorske reprezentacije reči i modeli zasnovani na transformer arhitekturi.

Ni jedan model iz prve grupe nije nadmašio transformer model, pa će dve grupe modela biti posmatrane odvojeno.

5.1. Modeli zasnovani na vektorskoj reprezentaciji reči

U tabeli 5.1.1 prikazani su F1 mera za obe klase i tačnost. Tačnost se kreće u rasponu od 82% do 86%, u zavisnosti od modela. Zanimljivo je da se naš osnovni model najkraće trenirao, a postigao je istu tačnost kao i najbolji model. Zaključujemo da produbljanjem mreže i dodavanjem potpuno povezanih slojeva nismo podigli performanse već samo produžili vreme treniranja.

Najbolji model iz ove grupe je model sa konvolutivnim i LSTM slojevima. Ovaj model postigao je F1 meru od 85% za sarkastične naslove, F1 meru od 87% za nesarkastične naslove i tačnost od 86%. Najbolji model je dostigao tačnost od 86%, što su iste performanse kao u [15].

Tabela 5.1.1. Metrika nad test skupom modela zasnovanih na vektorskoj reprezentaciji reči

Modeli	F1 Sarkastični naslovi	F1 Nesarkastični naslovi	Tačnost
Osnovni	0,84	0,87	0,86
Potpuno povezani slojevi	0,83	0,87	0,85
Konvolutivni i LSTM	0,85	0,87	0,86
LSTM i nepretrenirani embedding sloj	0,81	0,87	0,86
GRU i nepretrenirani embedding sloj	0,80	0,86	0,85
GRU, konvolutivni, potpuno povezani, pooling, dropout slojevi	0,71	0,79	0,82

5.2. Modeli zasnovani na transformer arhitekturi

U tabeli 5.2.1 su prikazani rezultati za transformer modele. Prikazani su F1 mera za obe klase i tačnost. Tačnost se kreće u rasponu od 90% do 94% u zavisnosti od modela.

Tabela 5.2.1. Metrika nad test skupom modela zasnovanih na transformer arhitekturi

Modeli	F1 Sarkastični naslovi	F1 Nesarkastični naslovi	Tačnost
BERT	0,91	0,93	0,92
BERT sa dodatnim klasifikacionim slojevima	0,91	0,93	0,92
roBERTa	0,92	0,94	0,93
roBERTa sa uvećanim <i>batch</i> -om i brojem warmup koraka	0,93	0,95	0,94
T5	0,88	0,91	0,90
GPT-2	0,89	0,91	0,90

BERT model je pokazatelj da transformer modeli postižu bolje rezultate od modela zasnovanih na vektorskoj reprezentaciji reči. Možemo primetiti da dodavanje klasifikacionih slojeva u dekoderskom delu, nije pomoglo pri klasifikaciji. Ovo je bilo očekivano jer se većina parametara modela nalazi u enkoderskom delu, pa dodavanje nekoliko slojeva u dekoderskom delu neće dovesti do primetnog poboljšanja performansi. Ovde, performanse nisu opale, ali je produženo vreme treniranja.

Najbolji model je bio roBERTa model sa uvećanim *batch*-om i uvećanim brojem *warmup* koraka u optimizacionoj funkciji. roBERTa model se najduže trenirao, ima najviše parametara i postigao je najbolju tačnost od 94%, najbolju F1 meru za sarkastične naslove od 93% i najbolju F1 meru za nesarkastične naslove od 95%. Poređenja radi, transformer model u radu [6] treniran i evaluiran nad više skupova podataka, uključujući i skup podataka koji smo mi koristili, postigao je F1 meru od 92% i tačnost od 92%.

T5 je koristio samo enkoderski deo i imao je znatno manje parametara u odnosu na BERT i roBERTa modele. Zbog toga je postigao lošije rezultate i tačnost od 90%.

GPT-2 je postigao rezultat vrlo sličan T5 modelu. S obzirom da je unidirekcion model, a ne bidirekcion model, rezultati su lošiji u odnosu na BERT i roBERTa modele.

6. ZAKLJUČAK

U ovom radu je rešavan problem detekcije sarkazma u naslovima novinskih članaka. Problem je rešavan pomoću modela zasnovanih na neuronskim mrežama koje koriste vektorske reprezentacije reči i modela zasnovanih na transformer arhitekturi. Isprobane su potpuno povezane neuronske mreže, mreže sa LSTM slojevima, sa GRU slojevima, u kombinaciji sa konvolutivnim slojem, ne pretreniranim i pretreniranim *Embedding* slojem. Njihova tačnost se kreće u rasponu od 82% do 86%. Ni jedan model iz prve grupe nije nadmašio transformer modele. Od modela zasnovanih na transformer arhitekturi isprobani su BERT, roBERTa, T5, GPT-2. Tačnost za transformer modele se kreće u rasponu od 90% do 94%. Najbolji rezultat je postigao pretrenirani roBERTa model sa tačnošću od 94%.

7. LITERATURA

- [1] Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- [2] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V., 2019. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

- [3] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. and Sutskever, I., 2019. Language models are unsupervised multitask learners. OpenAI blog, 1(8), p.9.
- [4] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W. and Liu, P.J., 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv preprint arXiv:1910.10683.
- [5] Misra, R. and Arora, P., 2019. Sarcasm detection using hybrid neural network. arXiv preprint arXiv:1908.07414.
- [6] Felbo, B., Mislove, A., Søgaard, A., Rahwan, I. and Lehmann, S., 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. arXiv preprint arXiv:1708.00524.
- [7] <https://twitter.com>
- [8] Akula, R. and Garibay, I., 2021. Interpretable Multi-Head Self-Attention Architecture for Sarcasm Detection in Social Media. *Entropy*, 23(4), p.394.
- [9] <https://reddit.com>
- [10] <https://keras.io/>
- [11] https://keras.io/keras_tuner/
- [12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- [13] <https://huggingface.co/>
- [14] Sun, C., Qiu, X., Xu, Y. and Huang, X., 2019, October. How to fine-tune bert for text classification?. In *China national conference on Chinese computational linguistics* (pp. 194-206). Springer, Cham.
- [15] Mandal, P.K. and Mahto, R., 2019. Deep CNN-LSTM with word embeddings for news headline sarcasm detection. In *16th International Conference on Information Technology-New Generations (ITNG 2019)* (pp. 495-498). Springer, Cham.

Kratka biografija:



Aleksandar Vujinović rođen je 1998. godine u Novom Sadu. Osnovne akademske studije završio je 2021. godine na Fakultetu tehničkih nauka, na kom brani master rad 2022. godine iz oblasti Softversko inženjerstvo i informacione tehnologije.