

Fake news challenge

Detekcija lažnih vijesti

Fran Mišić, Andrej Slapničar, Iva Sokolaj, Roko Torbarina

Travanj, 2022.

| | |
|---|---|
| Uvodni opis problema..... | 1 |
| Cilj i hipoteze problema | 2 |
| Pregled dosadašnjih istraživanja..... | 2 |
| Materijali, metodologija i plan istraživanja..... | 3 |
| Očekivani rezultati predloženog projekta..... | 4 |
| Popis literature..... | 4 |

Uvodni opis problema

Automatsko prepoznavanje istinitosti vijesti je izazovan problem te ima značajnu ulogu u političkom i socijalnom oblikovanju društva. *The New York Times* definirao je *Fake news* (lažne vijesti) kao izmišljene priče s namjenom da zavaravaju.¹ 2016. godine je organizirano natjecanje *Fake News Challenge* u kojem je zadatak pomoću umjetne inteligencije detektirati lažne vijesti. Zbog težine, natjecanje je podijeljeno u nekoliko dijelova. U prvom dijelu, zadatak je bio odrediti odnos naslova i sadržaja vijesti. Za treniranje modela trebalo je koristiti skup podataka² od 49972 instanci koje se sastoje od naslova vijesti, sadržaja vijesti te odnosa naslova i sadržaja koji može biti jedan od: slaganje, neslaganje, diskusija i nepovezanost.

¹ <https://www.nytimes.com/2016/12/06/us/fake-news-partisan-republican-democrat.html>

² <https://github.com/FakeNewsChallenge/fnc-1>

| | |
|--------------------------|---|
| Naslov vijesti | Explosion reported near the Nicaraguan capital attributed to meteorite impact, but experts, including NASA, cast doubt on claims |
| Sadržaj vijesti | News has been circulating about a potential meteorite strike near Managua, Nicaragua late Saturday night, just 13 hours or so before the close flyby of 20-m asteroid 2014 RC, leading some to suggest that the two events are related. : While this particular event is looking more and more like a false alarm with time, it should be noted that fireballs blaze through our skies every day, as tons of material is swept up by Earth as the planet orbits the Sun. Many of these are missed because they occur during the day, or over regions of the planet that aren't heavily populated. |
| Odnos naslova i sadržaja | agree |

Cilj i hipoteze problema

Cilj *Fake News Challenge*-a je istražiti kako umjetna inteligencija, posebno strojno učenje i obrada prirodnog jezika (NLP), može doprinijeti rješavanju problema lažnih vijesti.

Pregled dosadašnjih istraživanja

Natjecanju je pristupilo mnoštvo timova iz područja znanosti i industrije. Prva tri tima dobili su novčane nagrade od \$1000, \$600, \$400 te su njihova istraživanja i izvorni kod su javno dostupni kao *GitHub* repozitoriji:

1. SOLAT in the SWEN - <https://github.com/Cisco-Talos/fnc-1>
2. Athene (UKP Lab) - https://github.com/hanselowski/athene_system
3. UCL Machine Reading - <https://github.com/uclnlp/fakenewschallenge>

Prvi tim je koristio kombinaciju dubokih konvolucijskih neuronskih (*deep CNN*) i stabla odlučivanja bazirana na broju riječi, odnos naslova i sadržaja, *word2vec embeddings*-a³ i frekvenciju ključnih riječi (TF-IDF). Druga dva tima su koristili verziju neuronskih mreža *multilayer perceptron (MLP)* za rješavanje problema. U

³ <https://code.google.com/archive/p/word2vec/>

radovima su objasnili da je razlog zašto su njihove metode dale bolje rezultate od tehnika poput Bi-LSTM-ove ili CNN-ove u relativno malom broju jedinstvenih tekstova danih u izazovu kojih je oko 1600. Sva tri tima su postigli oko 82% na setu za testiranje. Za rješavanje problema također se koriste sljedeće metode i alati:

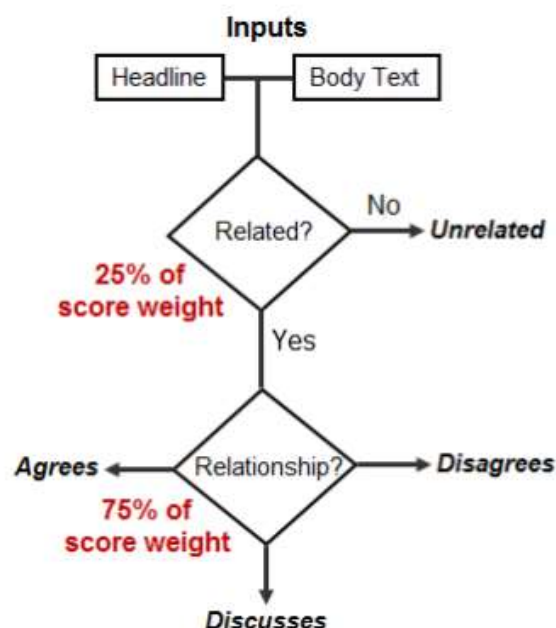
- SVM
- XGBoost
- K-Nearest Neighbors (KNN)
- Random Forests
- Decision Trees
- Neural Networks

Materijali, metodologija i plan istraživanja

Koristiti ćemo materijale s natjecanja za reproduciranje rezultata od najbolja tri tima i treniranje vlastitih modela te ispitivanje uspješnosti modela na podacima za testiranje.

Bodovanje uspješnosti modela:

- +0.25 ako je pogođena nepovezanost (*unrelated*)
- +0.25 ako je pogođeno da postoji povezanost (nije *unrelated*)
- +0.75 ako je pogođena točna povezanost (*agree*, *discuss* ili *disagree*)



Problemu ćemo pristupiti raznim algoritmima. Prvo ćemo proučiti jednostavnije modele poput *Decision Tree* ili KNN te ćemo probati iskoristiti *SoA textual entailment* modela za *feature extraction* i kreirati novi model na temelju dubokog učenja.

Očekivani rezultati predloženog projekta

Rezultati i broj bodova pojedinih ekipi FNC-1 natjecanja mogu se dohvatiti na: <https://competitions.codalab.org/competitions/16843#results>.

Naš cilj je reproduciranje algoritama i metoda koje su koristila prva tri tima te te pokušati ostvariti bolji uspjeh koristeći kombinaciju njihovih metoda i drugih metoda strojnog učenja.

Popis literature

1. <http://www.fakenewschallenge.org/>
2. <https://github.com/FakeNewsChallenge/fnc-1>
3. <https://arxiv.org/abs/1806.05180>