

ტექნიკური კომენტარების კლასიფიკაცია ქართული ონლაინ ფორუმის დისკუსიების
გამოყენებით

ნინელი ლაშქარაშვილი

ელ-ფოსტა: ninelilashkarashvili78@gmail.com

კომპიუტერული მეცნიერებების დეპარტამენტი
ზუსტ და საბუნებისმეტყველო მეცნიერებათა ფაკულტეტი
ივ. ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტი
უნივერსიტეტის ქ. 13, თბილისი

ხელმძღვანელი: მაგდა ცინცაძე
ფიზ.მათ მეცნიერებათა კანდიდატი
კომპიუტერული მეცნიერებების დეპარტამენტი

ტექნოლოგიების განვითარებასთან ერთად, მონაცემების ოდენობა საგრძნობლად მატულობს, მათ შორის ტექსტური ინფორმაციის გენერირება სულ უფრო ფართო მასშტაბიანი ხდება. ტექსტური მონაცემების დამუშავება ერთ-ერთი ყველაზე მნიშვნელოვანი საკითხია კვლევებსა და აპლიკაციებში. სენტიმენტების ანალიზი (ტექსტის კლასიფიკაცია) არის ერთ-ერთი გავრცელებული მეთოდოლოგია, რომელიც ტექსტურ ინფორმაციაში სუბიექტურ აზრსა და დამოკიდებულებას აანალიზებს. ეს ტექნიკა გამოიყენება მარკეტინგული სტრატეგიის გასაუმჯობესებლად, რეკომენდაციების შემუშავებისათვის, მომხმარებლის კმაყოფილების დონის გასაზრდელად (González-Fierro, 2017) და ა.შ. განსაკუთრებით აღსანიშნავია ამ მიდგომის აქტუალურობა სოციალური მედიის ტექსტური ინფორმაციის დამუშავებაში, რათა თავიდან იქნას აცილებული კიბერ-შეტევები, დამთრგუნველი თუ უცენზურო კომენტარები და კიბერბულინგი. განსაკუთრებით აქტუალურად Twitter-ისა და Facebook-ის მონაცემებს ანალიზებენ: 2014-2016 წლებში მომხმარებლის მიმოხილვებზე მეტად პოპულარული სოციალური მედიის პოსტების შინაარსის ანალიზი გახდა (Mäntylä et al, 2018).

პროექტში, ღრმა დასწავლის (DL) მოდელის გამოყენებით განხორციელდა ქართული ონლაინ საჯარო დისკუსიების პლატფორმის - თბილისი ფორუმის კომენტარების დამუშავება. პოლიტიკის სექციიდან ამოღებული 10000 კომენტარი შემდგომში გადამუშავდა და შინაარსის შესაბამისად თითოეული აღინიშნა, როგორც უცენზურო (toxic) ან არა-უცენზურო (non-toxic). შედეგად, 4639 კომენტარი აღმოჩნდა toxic, ხოლო 5361 non-toxic. შევნიშნავთ, რომ მომხმარებლის საიდენტიფიკაციო ინფორმაცია გამოყენებული არ ყოფილა.

ნაშრომში გამოყენებული ღრმა დასწავლის ალგორითმები იყენებენ Attention მექანიზმს (Vaswani et al, 2017) და biRNN (Schuster & Paliwal, 1997) მოდელზე დაფუძნებული.

ტოქსიკური კომენტარების კლასიფიკაციისთვის იმპლემენტირებული არქიტექტურებია: transformer encoder only, biRNN, biLSTM და biGRU. თითოეული არქიტექტურისთვის შეიქმნა 2 განსხვავებული მოდელი: პირველი fastText-ის ქართული სიტყვების წინასწარ გაწვრთნილი ჩადგმების (embeddings) გამოყენებით (Grave et al, 2018), ხოლო მეორე - მათ გარეშე: კერძოდ, ქართული ტექსტის დასამუშავებელი ფუნქციები იყო გამოყენებული, მიღებული შედეგი კი ჩადგმის შრეს(embedding layer) გადაეწოდა მონაცემებად.

იმისათვის, რომ toxic და non-toxic კომენტარების პროპორცია ერთნაირი ყოფილიყო საწვრთნელ (train), ვალიდაციისა (val) და სატესტო (test) მონაცემებში, გამოყენებული იყო stratified 5-fold cross-validation. ამასთანავე, ამ მეთოდით თითოეული მოდელი რამდენჯერმე გაეშვა მონაცემების შემთხვევითად არჩეულ განაწილებაზე train/val/test სიმრავლეებისათვის. საუკეთესო შედეგი transformer მოდელმა აჩვენა წინასწარ გაწვრთნილი ჩადგმების (pre-trained embeddings) გამოყენების გარეშე (0.869 +/- 0.009 ACC, 0.931 +/- 0.012 AUC).

ლიტერატურა

- González-Fierro, M. (2017). A Gentle Introduction To Text Classification And Sentiment Analysis. *Miguelgferro.com* <https://miguelgferro.com/blog/2017/a-gentle-introduction-to-text-classification-and-sentiment-analysis/>.
- Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning word vectors for 157 languages. arXiv preprint arXiv:1802.06893.
- Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep Learning--based Text Classification: A Comprehensive Review. *ACM Computing Surveys (CSUR)*, 54(3), 1-40.
- Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, 27, 16-32.
- Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11), 2673-2681.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762.