

Незважаючи на доволі не погані показники точності у групах «Reaction» і «Talk», 71% і 87% відповідно, графіки точності між тренувальними і тестовими наборами даних сильно відрізнялися, що піддає сумніву їхню доцільність у навчанні моделі. Проте, однією із причин, чому різниця точності аналізу цих груп не дуже висока – це фонетичні особливості котячого нявкоту [5]. Ці тварини створили власний метод спілкування із людьми, тому звуки, які вони видають, скоріш за все потребують врахування контексту прояву вокалізації.

Варто також зважати на те, що для аналізу використовувалася не велика база знань – лише з одного джерела, що спеціалізується на звукових ефектах. Тому, було вирішено проаналізувати MFCCs звуків завдяки T-розподіленому вкладенню стохастичної близькості (t-distributed Stochastic Neighbor Embedding, далі t-SNE) [6]. Після багатьох ітерацій алгоритму стало очевидним, що маркування даних на ресурсі є далеким від ідеального. Деякі категорії звуків не сформували видимі кластери. Були кластери сформувалися занадто близько один до одного. А ще були категорії, які сформували чіткі кластери, але містили дані, що були на занадто великій відстані від решти. З цього можна констатувати, що незважаючи на певну точність класифікації звуків, вони потребують додаткової або нової категоризації.

Висновки. Тренування моделей машинного навчання завжди потребує великої кількості даних. Якщо це якісь специфічні дані, що потребують довготривалого отримання, то виникає спокуса знайти їх у мережі Інтернет. Незважаючи на можливості цього ресурсу, треба враховувати і ризики його застосування. У випадку звуків тварин, я дійшов висновку, що краще займатися або власним видобутком даних або уважно застосовувати методи кластеризації отриманих з мережі записи.

Список використаної літератури

- [1]. Порівняльний аналіз робіт із дослідження поведінки тварин за їхніми звуками, А. Михайлів. (2024). Herald of Khmelnytskyi National University. Technical Sciences, 333(2), с. 370-375. <https://doi.org/10.31891/2307-5732-2024-333-2-57>.
- [2]. Soundly. The Complete Sound Effects Platform, Soundly AS, <https://getsoundly.com/>.
- [3]. Кероване навчання, Вікіпедія. Вільна енциклопедія, https://uk.wikipedia.org/wiki/Кероване_навчання.
- [4]. Cepstrum and MFCC, T. Bäckström, Aalto University Wiki, <https://wiki.aalto.fi/display/ITSP/Cepstrum+and+MFCC>.
- [5]. Paralinguistic information and biological codes in intra-and interspecific vocal communication: A pilot study of humans and domestic cats, S. Schötz. 2019.
- [6]. Introduction to T-Sne for High Dimensional Visualization, Digital Humanities Tools and Techniques II, Open Library, <https://ecampusontario.pressbooks.pub/nudh3/chapter/introduction-to-t-sne-for-high-dimensional-visualization/>.

УДК: 004.8

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ МЕТОДУ КЛАСИФІКАЦІЇ ТЕКСТІВ ЗА ВМІСТОМ ПРОПАГАНДИ НЕЙРОМЕРЕЖЕВИМИ МОДЕЛЯМИ ГЛИБОКОГО НАВЧАННЯ

Молчанова М.О. (m.o.molchanova@gmail.com)
Хмельницький національний університет (Україна)

Досліджено ефективність методу класифікації текстів за вмістом пропаганди нейромережевими моделями глибокого навчання, який дозволяє виявляти як явні, так і приховані пропагандистські повідомлення, що є критичним для ефективного розпізнавання пропаганди. Цей метод об'єднує традиційні рекурентні нейронні мережі з довготривалою пам'яттю та трансформери, що сприяє кращому розумінню послідовності та розумінню контексту текстового контенту. Розроблений метод дозволяє досягнути точності 97.83 % при застосуванні аугментації для навчальної вибірки, та 96.94% без застосування аугментації.

Пропаганда є невід'ємною складовою інформаційних маніпуляцій, яка включає різні форми, методи та засоби впливу на людей з метою зміни їхніх психологічних характеристик у бажаному напрямку. Тому своєчасне виявлення пропаганди є важливим напрямком наукових досліджень.

Метод виявлення пропаганди у тексті нейромережевими моделями глибокого навчання призначений для автоматизованої ідентифікації текстів, які містять пропагандистські елементи. Запропонований метод дозволяє виявляти як явні, так і приховані пропагандистські повідомлення, використовуючи поєднання традиційних рекурентних нейронних мереж з довготривалою пам'яттю та трансформерів [1, 2]. Також застосовується механізм аугментації навчальних текстових даних, що дозволяє збільшити кількість навчальних прикладів [2]. Схема запропонованого методу наведена на рисунку 1.

Метод працює шляхом трансформації вхідних текстових даних через навчену нейромережеву модель глибокого навчання у вихідні дані у вигляді відсоткової оцінки наявності пропаганди у тексті та присвоєння одного із трьох класів: «текст без пропаганди», «пропагандистський текст» або ж «підозрілий текст».

Етап 1 включає попередню обробку тексту для класифікації, що передбачає перетворення тексту в нижній регістр, видалення стоп-слів та пунктуації тощо. Після попередньої обробки текст перетворюється у числові послідовності, які подаються на вхід навченій нейромережевій моделі глибокого навчання з гібридною архітектурою для подальшої оцінки відсоткової ймовірності наявності пропаганди.

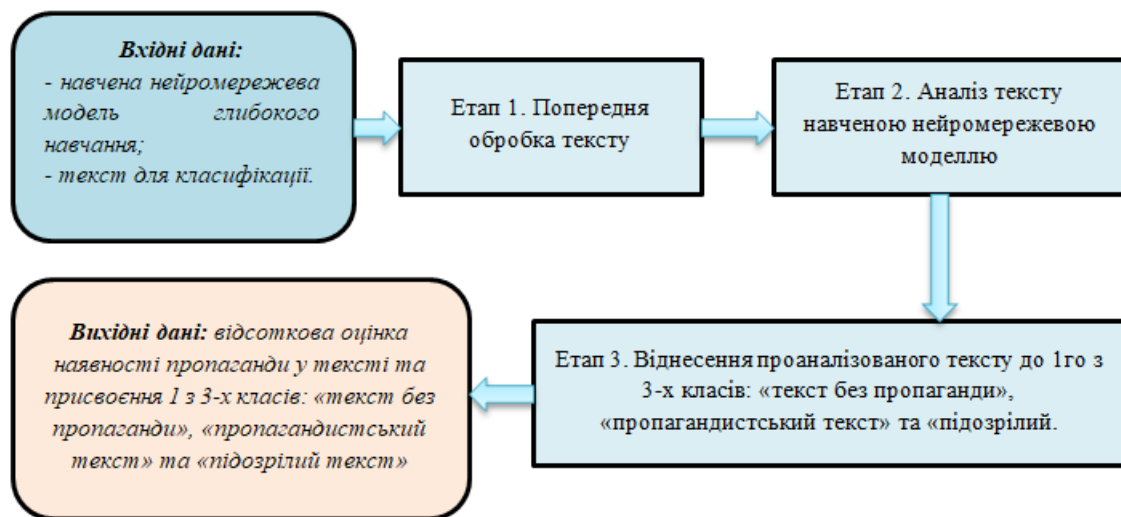


Рисунок 1 – Метод виявлення пропаганди у тексті нейромережевими моделями глибокого навчання

Етап 2 полягає в аналізі тексту навченою нейронною моделлю, яка здійснює числову оцінку рівня наявності пропаганди у тексті. Для цього застосовується нейронна мережа глибокого навчання з гібридною архітектурою, яка поєднує традиційну рекурентну нейромережу з довготривалою пам'яттю і трансформери. Це дозволяє досягти більш глибокого розуміння послідовності та контексту текстового контенту.

Етап 3 передбачає віднесення проаналізованого тексту до одного з трьох класів: «текст без пропаганди», «пропагандистський текст» або «підозрілий текст». Для цього емпіричним шляхом були визначені межі для кожного класу: «текст без пропаганди» відповідає значенням від 0 до 0.45, «підозрілий текст» має межі від 0.45 до 0.55, а «пропагандистський текст» - від 0.55 до 1. Ці межі можуть змінюватися та налаштовуватися в залежності від видів пропаганди та специфіки користувацьких даних.

Вихідними даними є відсоткова оцінка наявності пропаганди у тексті та його класифікація як «текст без пропаганди», «пропагандистський текст» або «підозрілий текст».

За розробленим методом було створено програмну реалізацію засобами мови Python. Інтерфейс застосунку наведено на рисунку 2.

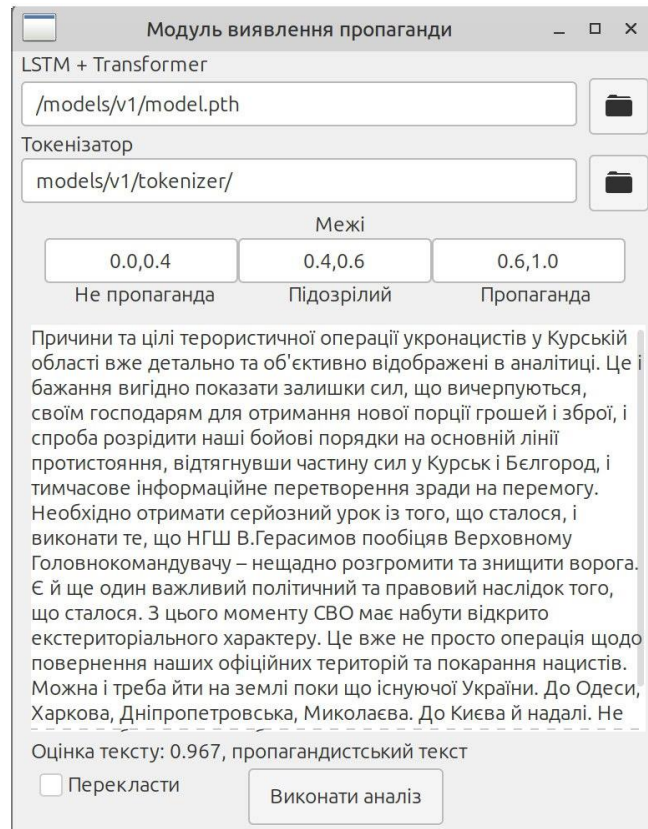


Рисунок 2 – Інтерфейс застосунку для реалізації методу

За допомогою створеного програмного забезпечення, що реалізує метод виявлення пропаганди у тексті нейромережевими моделями глибокого навчання було досліджено його ефективність. Вдалося досягти значення 96,94 % за метрикою Ассигасу без використання аугментації навчальної вибірки та значення 97.83 % з використанням аугментації для навчальної вибірки. Графік проведених експериментів наведений на рисунку 3.

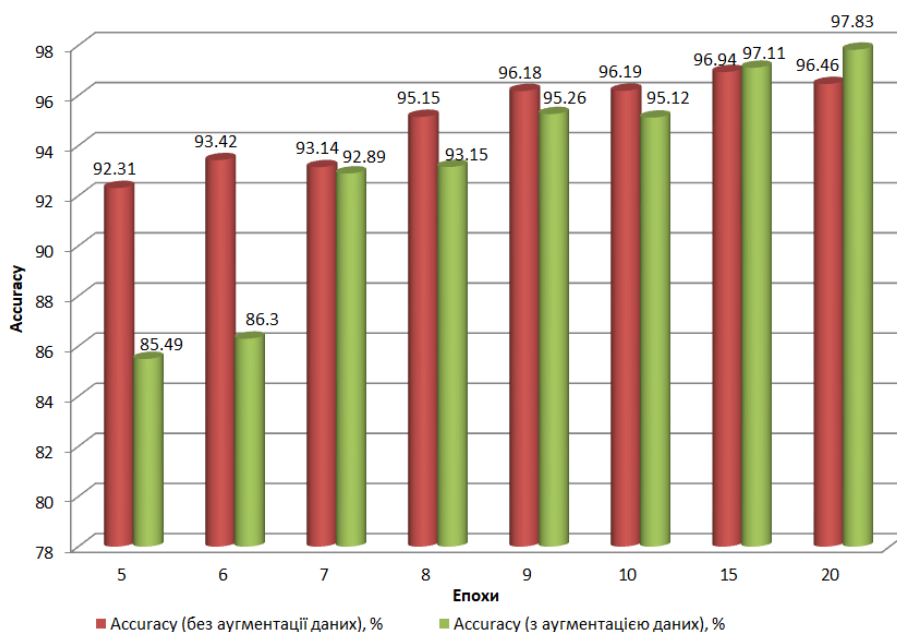


Рисунок 3 – Порівняння метрики Ассигасу із застосуванням аугментації та без неї

Як видно з рисунку 3, аугментація дозволяє покращити результат виявлення пропаганди на 0.89 %, що пояснюється розширенням навчальної вибірки та можливістю навчити нейромережу

прихованим залежностям. Без аугментації вдалось отримати показник метрики Accuracy на рівні 96.94%.

Отже, результати свідчать про здатність запропонованого методу виявлення пропаганди у тексті нейромережевими моделями глибокого навчання ефективно класифікувати тексти зі вмістом пропаганди. Введення додаткової категорії «підозрілий текст» дозволило підвищити показники Precision та Recall, що, в свою чергу, забезпечило можливість автоматизованої модерації текстів на предмет пропаганди з помилкою не більше 1.83% для хибного виявлення пропаганди.

Список використаних джерел

[1] Zalutska O., Molchanova M., Sobko O., Mazurets O., Pasichnyk O., Barmak O., Krak I. Method for Sentiment Analysis of Ukrainian-Language Reviews in E-Commerce Using RoBERTa Neural Network. CEUR Workshop Proceedings, 2023, vol. 3387, pp. 344–356.

[2] Молчанова М.О., Залуцька О.О., Бармак О.В. Метод інтелектуального аналізу тональності текстів. Матеріали XII Всеукраїнської науково-практичної конференції «Глушковські читання». Київ – 2023. с. 113-116.

[3] Krak I., Zalutska O., Molchanova M., Mazurets O., Manziuk E., Barmak O. Method for Neural Network Detecting Propaganda Techniques by Markers With Visual Analytic. CEUR Workshop Proceedings, 2024, vol. 3790, pp. 158-170.

УДК 004.896

РОЗРОБКА ПРОТЕЗНОГО СЕРВІСУ

Наймитенко С. І., Подорожняк А. О. (s.naimytenko@gmail.com,
andrii.podorozhniak@khp.edu.ua)

Національний технічний університет
“Харківський політехнічний інститут”(Україна)

В тезах розглядається актуальність, проблематика та рішення, що до розвитку протезування верхніх кінцівок. Аналізується ситуація на світовому та українському ринку протезування де приводяться відомі приклади протезування, та пропонується власний шлях розробки сервісу. У висновку розглядається результат розробки та розглядається варіанти покращень сервісу.

Важливість протезування у військовій реалії нашої країни це факт котрий не можна заперечувати. Наразі індустрія протезування кінцівок в нашій країні робить перші кроки, але робить це доволі впевнено [1]. Прикладом є компанія Esper Bionics, що наразі отримали близько 3-х мільйонів доларів інвестицій, що є доволі гарним показником враховуючи, що одна з провідних компаній Open Bionics має 8.9 мільйонів доларів інвестицій [2]. На жаль це все одно не достатня сума для провідних досліджень та розробок, що, вивести цей стартап заснований у 2019-му році на світову першість. Та за інноваційними розробками, як зазначалось раніше стоять великі суми, що є не завжди можливими для середнього громадянина нашої країни. Для цього авторами було запропоновано ідею розробки саме бюджетного варіанту протеза верхньої кінцівки для забезпечення необхідних задач для життєдіяльності людини [3].

Для того що б розуміти ситуація на ринку протезування було розглянуто звіти аналітичних компаній що прогнозують активне зростання індустрії в подальші 10 років. Таких висновків вийшли аналітичні компанії Strategic Market Research, Allied Market Research [4, 5]. Причиною цього є розвиток технологій, збільшення захворюваності у людей та збільшення кількості лікарів.

Аналізуючи типи протезів було переглянути історію їх розвитку, де спочатку протези являли собою косметичний інструмент, що надав мінімум функціоналу, але надав відчуття присутності кінцівки. Згодом це переросло в повноцінний високотехнологічний інструмент де наразі свою перевагу тримають саме міоелектричні протези, які функціонують внаслідок напруження м'язів