УДК 004.056.5:004.73

DOI: 10.31673/2412-9070.2022.061416

А. В. ЛЕМЕШКО, доктор філософії;
К. Ю. ШУЛЬЖЕНКО, студентка;
А. Ю. БЕРЕЗОВСЬКИЙ, студент;
В. С. ГАЛЕТА, студент,
Державний університет телекомунікацій, Київ

ФІЛЬТРАЦІЯ ЗАБОРОНЕНОГО КОНТЕНТУ

Актуальність цього дослідження полягає в необхідності фільтрації контенту з високою точністю через створення оптимальних варіацій архітектур нейронних мереж. Наявні сьогодні рішення дають низьку точність фільтрації, що призводить до блокування потенційно безпечного контенту. Повна відсутність фільтрації призведе до того, що неповнолітні користувачі та інші вразливі групи отримають до нього доступ, що неприпустимо.

Об'єктом дослідження є процеси фільтрації контенту.

Предметом дослідження є методи і технології побудови нейронних мереж для фільтрації контенту.

Мета статті полягає в підвищенні точності фільтрації контенту за рахунок розроблення архітектури нейронної мережі.

Як засіб розроблення системи було вибрано PyCharm та мову програмування Python. Інструментами розроблення були вибрані бібліотеки SciPy та Keras, бібліотека для імпорту та експорту даних Pickle та бібліотека для роботи з алгоритмами нейронних мереж Sckit-learn.

Результатом роботи є побудована архітектура нейронної мережі для фільтрації забороненого контенту.

Ключові слова: нейронна мережа; заборонений контент; фільтрація контенту; архітектура нейронних мереж.

ВСТУП

Постановка проблеми. У сучасному світі все активніше набирає обертів процес діджиталізації та інтегрування все більшої кількості програмних продуктів у життя рядового члена соціуму. Кожного дня з'являються нові застосунки, які допомагають людям генерувати терабайти контенту, зокрема TikTok, Instagram, Telegram, Linked In тощо.

Проте зі збільшенням кількості контенту зростає і кількість матеріалів, що порушують правила платформи або закони країни, громадянином якої є користувач.

Кожного дня в соціальних мережах з'являються фото і відеоматеріали, на яких присутні паління, алкоголь, зброя, пропаганда наркотиків тощо. І саме в цей момент користувачам на допомогу приходить явище фільрації контенту.

Аналіз останніх досліджень і публікацій. Було проведено аналітичний огляд методів фільтрації контенту. Кожний з методів має власні переваги і недоліки. У ході аналізу було визначено, що найефективнішим варіантом на поточний момент є використання нейронних мереж, які навчаються на прикладах забороненого контенту для його подальшої фільтрації.

Мета статті — аналіз актуальності використання нейронних мереж для фільтрації забороненого контенту та явища фільтрації контенту в цілому.

ОСНОВНА ЧАСТИНА

Інтернет-фільтр — це програмне забезпечення, яке обмежує або контролює вміст, до якого може отримати доступ користувач інтернету, особливо коли він використовується для обмеження матеріалів, що доставляються через Інтернет, електронну пошту чи іншими засобами. Програмне забезпечення для контролю вмісту визначає, який вміст буде доступним або заблокованим [1].

Такі обмеження можуть застосовуватися на різних рівнях: уряд може спробувати застосувати їх по всій країні, або вони можуть, наприклад, застосовуватися постачальником послуг інтернету до своїх клієнтів, роботодавцем до свого персоналу, школою для своїх учнів, бібліотекою для відвідувачів, батьками до комп'ютера дитини або окремими користувачами до власних комп'ютерів.

Фільтри можуть бути реалізовані багатьма різними способами: за допомогою програмного забезпечення на персональному комп'ютері, через мережну інфраструктуру, таку як проксі-сервери, DNS-сервери або брандмауери, які забезпечують доступ до інтернету. Жодне вирішення не забезпечує повного охоплення, тому більшість компаній розгортають поєднання технологій для досягнення належного контролю вмісту відповідно до своєї політики.

Сьогодні одним з основних методів фільтрації контенту є фільтрація контенту за допомогою нейронних мереж [2].

Браузерні фільтри. Рішення для фільтрації вмісту на основі вебпереглядача є найпростішим рішенням для фільтрування вмісту та реалізується через стороннє розширення для браузера.

Фільтри електронної пошти. Фільтри електронної пошти діють на інформацію, що міститься в тілі листа, у заголовках листів, як-от відправник і тема, а також на вкладення електронної пошти,

© А. В. Лемешко, К. Ю. Шульженко, А. Ю. Березовський, В. С. Галета, 2022

щоб класифікувати, приймати чи відхиляти повідомлення. Зазвичай використовуються фільтри Байєса, тип статистичного фільтра. Доступні як клієнтські, так і серверні фільтри.

Клієнтські фільтри. Цей тип фільтра встановлюється як програмне забезпечення на кожному комп'ютері, де потрібна фільтрація. Зазвичай цим фільтром може керувати, вимикати чи видаляти кожен, хто має права адміністратора в системі. Клієнтський фільтр на основі DNS мав би налаштовувати DNS Sinkhole, наприклад Pi-Hole.

Інтернет-провайдери з обмеженим (або відфільтрованим) вмістом. Інтернет-провайдери з обмеженим (або фільтрованим) вмістом — це постачальники Інтернет-послуг, які пропонують доступ лише до певної частини Інтернет-вмісту за згодою або в обов'язковому порядку. Кожен, хто підписався на цей тип послуг, підпадає під обмеження. Тип фільтрів може бути використаний для здійснення державного регуляторного або батьківського контролю над абонентами.

Мережна фільтрація. Цей тип фільтра реалізується на транспортному рівні як прозорий проксі або на прикладному рівні як вебпроксі. Програмне забезпечення для фільтрації може містити функцію запобігання втраті даних для фільтрації вихідної та вхідної інформації. Усі користувачі підлягають політиці доступу, визначеній установою. Фільтрування можна налаштувати, тому бібліотека середньої школи шкільного округу може мати інший профіль фільтрації, ніж бібліотека молодшої школи округу.

Фільтрація на основі DNS. Цей тип фільтрації реалізовано на рівні DNS і намагається запобігти пошуку доменів, які не відповідають набору політик (або батьківського контролю, або правил компанії). Кілька безкоштовних публічних служб DNS пропонують параметри фільтрації як частину своїх послуг. DNS Sinkholes, наприклад Pi-Hole, також можна використовувати для цієї мети, але лише на боці клієнта.

Фільтри пошукових систем. Багато пошукових систем, таких як Google і Bing, пропонують користувачам можливість увімкнути фільтр безпеки [3]. Коли цей фільтр безпеки активовано, він відфільтровує невідповідні посилання з усіх результатів пошуку. Якщо користувачі знають фактичну URL-адресу вебсайту, який містить відвертий або дорослий вміст, вони мають можливість отримати доступ до цього вмісту без використання пошукової системи. Деякі провайдери пропонують орієнтовані на дітей версії своїх двигунів, які дозволяють лише веб-сайти, орієнтовані на дітей [4].

У процесі аналізу предметної сфери було розглянуто різні варіація наявних рішень, які надають послуги фільтрації забороненого контенту. Окремо слід виокремити алгоритми штучного інтелекту платформи TikTok. Ця платформа забезпечує блокування повного спектра забороненого контенту, починаючи від жорстокості й алкоголю, закінчуючи зброєю і сексуальним контентом. Водночас схожу систему фільтрації має мережа Instagram.

Проте, в обох мережах існують недоліки.

Мережа Instagram занадто лояльна і за наявності сумнівів вона, скоріше, не заблокує потенційно неприйнятний контент, зачасту такі алгоритми виявляють у нормальному режимі тільки жорстокість.

У алгоритмів TikTok ситуація прямо протилежна. За наявності найменших збіжностей зі списком забороненого контенту — система образу блокує контент, проте дуже часто відбуваються ситуації, в яких звичайну пляшку води платформа розпізнає як алкоголь, а складені два пальці — як зброю.

ВИСНОВКИ

З огляду на зазначене можна зробити висновок щодо високої актуальності створення різного роду інструментів для автоматичної фільтрації контенту з адекватною моделлю поведінки.

Список використаної літератури

1. ***Заяць В. М., Камінський Р. М.*** *Методи розпізнавання образів: навч. посіб. для студ. Нац. ун-т «Львів. політехніка». Львів, 2004. 173 с.*

2. ***Дэвід А. Форсайт, Джин Понс.*** *Computer Vision: A Modern Approach. M.: Вільямс, 2004. 928 с.*

3. ***Стокман Дж., Шапіро Л.*** *Computer Vision. M.: Біном. Лаборатория знаний, 2006. 752 с.*

4. ***Вапнік В. Н., Червоненкіс А. Я.*** *Теорія розпізнавання образів. М.: Наука, 1974. 416 с.*

A. Lemeshko, K. Shulzhenko, A. Berezovskyi, V. Haleta

FILTERING FORBIDDEN CONTENT

The relevance of this study lies in the need to filter content with high accuracy due to the creation of optimal variations of neural network architectures. The solutions available today provide low filtering accuracy, which leads to the blocking of potentially safe content. A complete lack of filtering will lead to the fact that minor users and other vulnerable groups will gain access to it, which is unacceptable.

In today's world, the process of digitization and integration of more and more software products into the life of an ordinary member of society is gaining momentum.

Every day there are new apps that help people generate terabytes of content like TikTok, Instagram, Telegram, Linked In, etc.

However, with the increase in the amount of content, the number of materials that violate the rules of the platform or the laws of the country of which the user is a citizen also increases.

Every day there are photos and videos on social networks that contain smoking, alcohol, weapons, drug propaganda, direct or indirect, etc.

At this moment, the phenomenon of content filtering comes to the rescue of users.

The object of research is content filtering processes.

The subject of research is the methods and technologies of building neural networks for content filtering.

The purpose of the work is to increase the accuracy of content filtering by developing a neural network architecture.

PyCharm and the Python programming language were chosen as a system development tool. The SciPy and Keras libraries, the Pickle data import and export library, and the Sckit-learn library for working with neural network algorithms were chosen as development tools.

The result of the work is the constructed neural network architecture for filtering prohibited content.

An analytical review of content filtering methods was conducted. Each of the methods has its own advantages and disadvantages. During the analysis, it was determined that the most effective option at the moment is the use of neural networks that learn from examples of prohibited content for its further filtering

An Internet filter is software that limits or controls the content that an Internet user can access, especially when it is used to restrict material delivered over the Internet via the Internet, email, or other means. Content control software determines what content is allowed or blocked.

No single solution provides complete coverage, so most companies deploy a mix of technologies to achieve adequate content control according to their policies.

Such restrictions can be applied at different levels: a government can try to apply them nationwide, or they can, for example, be applied by an Internet service provider to its customers, an employer to its staff, a school to its students, a library to visitors, a parent to a computer child or individual users to their own computers.

The purpose of the article is to analyze the relevance of using neural networks for filtering prohibited content and the phenomenon of content filtering in general.

Based on all of the above, we can conclude that it is highly relevant to create various tools for automatic content filtering with an adequate behavior model.

Keywords: neural network; prohibited content; content filtering; neural network architecture.