
DOI: <https://doi.org/10.15407/kvt203.01.026>

УДК 004.89

АНИСИМОВ А.В., д-р. фіз.-мат. наук,
член.-кор. НАН України, декан факультету
комп'ютерних наук та кібернетики
ORCID: 0000-0002-1467-2006
e-mail: anatoly.v.anisimov@gmail.com

БЕВЗА М.В., аспірант
ORCID: 0000-0002-2697-4968
e-mail: maksymbevza@gmail.co

БОБИЛЬ Б.В., аспірант,
ORCID: 0000-0002-9612-1071
e-mail: bobybohdan@gmail.com

Київський національний університет імені Тараса Шевченка,
вул. Володимирська, 60, м. Київ, 01033, Україна

ПРОГНОЗУВАННЯ ВІДГУКІВ НА ВІЗУАЛЬНО-ТЕКСТОВИЙ КОНТЕНТ З ВИКОРИСТАННЯМ НЕЙРОННИХ МЕРЕЖ

Вступ. Соціальні мережі надають можливість одержувати високо персоналізований досвід для їхніх користувачів, даючи їм змогу підписуватись на сторінки інших користувачів, які публікують релевантний і цікавий контент. Відгук на візуальний і текстовий контент повідомлення одержується від підписників у вигляді поширень, лайків та коментарів. Прогнозування реакції аудиторії на те чи інше повідомлення цих мереж стає чимдалі важливим завданням, оскільки публікація повідомлень носить незворотний характер і перед публікацією необхідно прогнозувати реакцію підписників.

Метою роботи є побудова системи, яка може прогнозувати реакцію аудиторії на чергову публікацію, враховуючи особливість кожної сторінки, унікальність її аудиторії та варіативність можливих реакцій.

Результати. Проаналізовано структуру контенту і реакцій у соціальних мережах та визначено особливості публікацій, які слід враховувати під час аналізу популярності та відгуку аудиторії на публікацію. Описано процес тренування нейронної мережі, що дає можливість натренувати нейронну мережу для кожної сторінки з урахуванням особливостей аудиторії. Характеристики текстової частини контенту попередньо сформовано за допомогою нейронної мережі BERT, а характеристики візуальної частини контенту — нейронної мережі VGG-16. Створено модель класифікації для трьох груп показників: лайків, поширень та емоційної направленості коментарів. Розроблено алгоритм передбачення реакції аудиторії на повідомлення, який об'єднує роботу компонентів нейронної мережі для екстракції візуально-семантичних ознак зображення, текстово-семантичних ознак текстових даних та

© АНИСИМОВ А.В., БЕВЗА М.В., БОБИЛЬ Б.В., 2021

мета-інформації досліджуваного повідомлення. Розроблено систему, яка опрацьовує у поєднанні візуальну і текстову частину контенту і таким чином дає можливість отримати повний контекст публікації. Аналіз роботи розробленої системи показав, що найкраще прогнозовано сентимент коментарів аудиторії, якість прогнозування лайків і репостів була децю нижчою.

Висновки. Використання розробленої нейронної мережі, яка складається з трьох повноз'язних шарів і на вхід отримує характеристики текстової та візуальної частин контенту та мета-дані про досліджувану сторінку, дає змогу комплексно оцінити реакцію користувачів на чергову публікацію конкретної сторінки в соціальній мережі.

Ключові слова: штучний інтелект, оброблення природної мови, комп'ютерний зір, соціальні мережі, прогнозування реакції аудиторії, текстовий та візуальний контент.

ВСТУП

Соціальні мережі надають можливість одержувати високоперсоналізований досвід для їхніх користувачів, даючи їм змогу підписуватись на сторінки інших користувачів, які публікують релевантний і цікавий контент.

Комерціалізація цих мереж теж відбулась за лічені роки після їхнього створення. Рекламу в цих мережах є дуже популярною. Рекламою також займаються користувачі, що ведуть власні сторінки і мають велику кількість підписників, які регулярно отримують нові повідомлення. Прогнозування реакції аудиторії на те чи інше повідомлення у цих мережах стає чимдалі важливим завданням, оскільки публікація повідомлень носить незворотний характер і перед публікацією необхідно прогнозувати реакцію підписників.

Попередні дослідження з прогнозування реакції аудиторії орієнтовано на великі веб-ресурси і не беруть до уваги контекст, за яким здійснюється прогнозування реакції.

ПОСТАНОВКА ПРОБЛЕМИ

Реакція аудиторії зазвичай розглядається в контексті різних жанрів і способу побудови історії та тексту для отримання різних реакцій [1]. Відзначено необхідність визначення, як саме спосіб викладення контенту впливає на те, чи прочитає аудиторія контент до кінця і, відповідно, чи будуть реакції на нього. Таку залежність між способом викладення інформації і майбутньою активністю читачів проаналізовано у [1] та зазначено значне підвищення цієї активності, якщо учасники мають справу з контентом у жанрі розповідання історій.

У дослідженнях емоційного окрасу реакцій користувачів соціальних мереж на події широкого масштабу (державного і глобального) є корисним використовувати не тільки бінарну позитивно-негативну, а й детальну класифікацію цих реакцій. Проведений аналіз афективних виразів у Twitter, зібраних авторами у Німеччині на основі схеми кодування [2], надав змогу зрозуміти спосіб сприйняття користувачами соціальних мереж загальної ситуації під час інциденту забруднення харчових продуктів ЕНЕС 2011 р. Аналіз їхньої детальнішої реакції допомагає визначати загальні настрої в суспільстві та оцінювати цілі, які постають перед людьми у соціумі, і мотиви, якими вони керуються.

Треба зазначити дослідження немаркованих даних для створення класифікатора настроїв Twitter за допомогою мереж Convolutional Neural Networks (CNN) та Long Short Term Memory (LSTM) [3]. Авторами використано метод оцінювання сентименту для повідомлень у соціальній мережі

Twitter за допомогою моделі згорткових та рекурентних нейронних моделей. Використано дві категорії сентименту — позитивний і негативний. Ця робота описує state-of-the-art з сентимент аналізу повідомлень.

Дуже цікаве завдання ще не знайшло свого розв’язання, а саме прогнозування реакції користувачів у вигляді «лайк» та «поширення» для конкретного повідомлення на сторінці користувача з певною тематикою.

Сучасні методи оброблення текстових даних базуються на застосуванні рекурентних нейронних мереж (RNN), а також системи Transformer [4].

Рекурентна нейронна мережа (RNN) може отримувати на вхід наперед невизначену кількість вхідних даних у послідовності і опрацьовувати їх послідовно. Щоправда, є практичні обмеження щодо кількості вхідних даних в одній послідовності, що зумовлено обмеженим розміром прихованого шару мережі. Рекурентні нейронні мережі мають велику популярність у сучасних методах оброблення текстів за рахунок їхньої можливості «пам’ятати» контекст речення у разі подання їм на вхід послідовності слів. Найпоширеніші реалізації рекурентних нейронних мереж є системи Long Short-Term Memory (LSTM) [3].

Ще один з підходів навчання нейронних мереж використано в мережі Transformer [4]. Ця нейронна мережа складається з багатьох повнозв’язних шарів, а також має вбудовану систему Attention, яка дає можливість нейронній мережі фокусуватися на найважливіших частинах речення на кожному рівні нейронної мережі.

Отже, наведені дослідження сконцентровано на аналізі окремих характеристик візуально-текстової інформації, таких як емоційне навантаження тексту, категоризація за темами та/або жанрами, класифікація зображень тощо.

Мета роботи — розробити систему, яка дасть можливість для будь-якої сторінки веб-ресурсу з великою популярністю (більше 5000 підписників) передбачити відгук читачів на чергове повідомлення. Наша робота має фокус на підлаштування під кожну окрему сторінку і на врахування деталей та попередніх коментарів саме з цієї сторінки, щоб максимально точно передбачити реакцію аудиторії конкретної сторінки.

ПІДГОТОВКА ВХІДНИХ ДАНИХ ДЛЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ РЕАКЦІЇ АУДИТОРІЇ НА ПОВІДОМЛЕННЯ НА ПЕРСОНАЛЬНІЙ СТОРІНЦІ КОРИСТУВАЧА

Дані для тренування. Збір даних для тренування накладає значні обмеження на процес тренування. Автоматичне завантаження даних користувачів, які не дали на це безпосередньої згоди, не є дозволеним. Тому дані для тренування та валідації було завантажено у напівавтоматичному режимі з даних авторів цієї статті та обмеженого кола осіб, що дали свою згоду на вивантаження їхніх даних.

Структура об’єкту дослідження — контенту і реакцій у соціальних мережах. У контексті соціальних медіа, таких як “Facebook” та “Instagram” публікований контент може складатися з трьох компонентів:

- візуальний контент — набір зображень або відео;
- аудіо контент;
- текстовий контент — текстові публікації та коментарі користувачів.



Рис. 1 Схематичне подання опублікованого контенту і реакції аудиторії на різних соціальних платформах (Facebook, Instagram)

Також соціальні платформи підтримують можливість для аудиторії виражати реакцію та ставлення до опублікованого контенту. Наразі способи вираження реакції аудиторії до контенту є стандартними практично для всіх соціальних платформ і можна виділити такі інструменти:

- лайк — інструмент для швидкого вираження позитивного ставлення до контенту, що зазвичай реалізовано у вигляді кнопки з лічильником;
- дизлайк — інструмент для швидкого вираження негативного ставлення до контенту, що зазвичай реалізовано у вигляді кнопки з лічильником;
- текстовий коментар;
- поширення (репост) — інструмент, який дає змогу аудиторії копіювати контент, публікувати і поширювати його на власних сторінках.

Схематично структуру опублікованих у соціальних мережах даних та реакції аудиторії надано на рис. 1.

Отже, аудиторія має багато способів виразити своє ставлення до опублікованого контенту. У схемі (Рис. 1) наведені найпоширеніші інструменти, проте варто враховувати, що будь-яка соціальна платформа може мати свої специфічні засоби для вираження ставлення аудиторії до інформації, що публікується.

Прогнозування популярності та відгуку аудиторії на публікацію є складною та комплексною задачею у загальному випадку за таких причин:

1) публікація складається зі слабоструктурованої інформації, яка має складну природу;

2) структура публікації може відрізнятися за різних платформ і може як включати, так і не включати ті чи інші засоби; наприклад можливість ставити дизлайки на деяких платформах відсутня, коментарі до публікації можуть бути відключеними автором тощо;

3) модерація контенту може бути різною, тобто автори можуть видаляти коментарі, деякі платформи підтримують можливість налаштування відображення публікації тільки для певної частини аудиторії тощо;

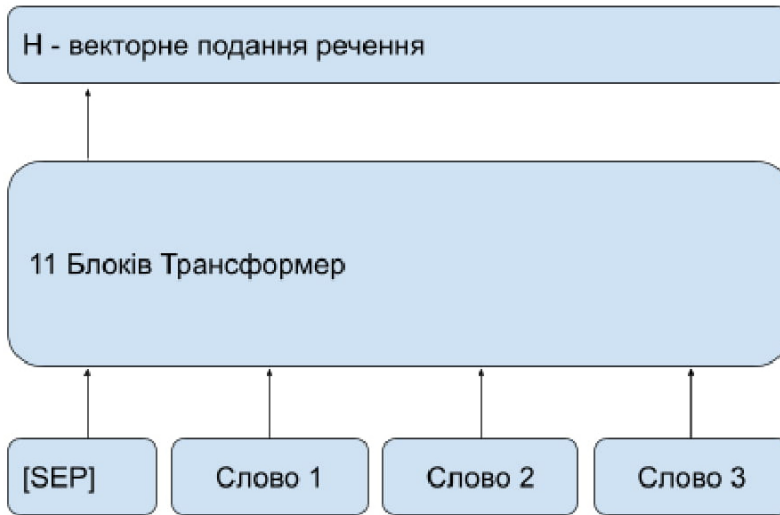


Рис 2. Схема роботи нейронної мережі BERT

4) аналіз кожної частини публікації (текст, зображення, відео) є тривіальним завданням.

У цій роботі пропонується система для аналізу візуально-текстового контенту публікації, що базується на нейронних мережах та основна ідея якої — це побудова швидкої для тренування нейронної мережі, специфічної для кожного автора, яка буде прогнозувати такі значення для нової публікації: кількості лайків, кількості поширень та емоційну направленість коментарів аудиторії (позитивна/негативна).

Отримання характеристик з текстової частини контенту. Для генерації ознак з текстових даних було використано нейронну мережу, побудовану на основі архітектури Transformer [4]. Оскільки навчання нейронної мережі для оброблення потребує величезних обсягів даних, то цю систему ми тренували на основі попередньо навченої мережі на мовних моделях системи BERT [5].

На рис. 2 зображено схему нейронної мережі BERT. Ця система дає змогу приймати на вході ціле речення, розраховувати 12-рівневу нейронно-мережеву архітектуру і отримати контекстно-залежні репрезентації слів. Цей підхід має перевагу над підходом незалежного оброблення слів, оскільки враховує контекст, в якому кожне слово використовується.

Для розв'язання цього завдання використано тільки перші 11 рівнів нейронної мережі задля того, щоб отримати репрезентацію слів з усіма контекстними ознаками, але без ознак, які потрібні для обчислення ймовірності появи слова у мовній моделі.

Кожне слово має 768-розмірний вектор, який відповідає його числовій репрезентації. Для побудови єдиного значення для всього тексту здійснено усереднення за всіма словами в тексті як спосіб узагальнення.

Отримання характеристик з візуальної частини контенту. Для генерації ознак із зображення використано згорткову нейронну мережу VGG-16 [6], а саме частину мережі зі згортковими шарами, які відповідають за екстракцію

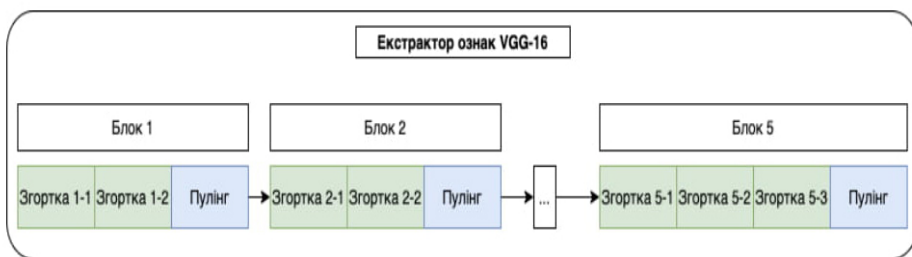


Рис. 3. Структура згорткових шарів мережі VGG-16

ознак, що описують вміст зображення. Вибір такої архітектури зумовлено тим, що її тренування було здійснено на наборі даних ImageNet [7] і ваги мережі для цього набору даних знаходяться у відкритому доступі. Альтернативний спосіб ініціалізації ваг мережі — це ініціалізація випадковими значеннями, що мають нормальний розподіл, параметри якого є математичним сподіванням та дисперсією, залежать від кількості вхідних значень та кількості вихідних значень шару [8]. Структуру згорткових шарів надано на рис. 3. Ця мережа дає наразі найкращі результати на наборі даних ImageNet, але простота її структури і ознаки, вивчені в процесі тренування на ImageNet, дають змогу швидко адаптувати її до інших завдань та дотреновувати за короткий час, що є значною перевагою для нашого завдання.

Нормалізація зображення здійснюється таким чином. Перед тим, як подати зображення в мережу, його розмір зменшується до 224x224 пікселів (стандартний розмір для задач класифікації зображень). Ця трансформація робиться для того, щоб пришвидшити оброблення даних та зменшити використання пам'яті. Після цього здійснюється нормалізація значень зображення — від нього віднімається середнє значення за кожним кольоровим каналом та ділиться на 255 — всі трансформації робляться по-піксельно. Після цих дій, зображення подається в мережу.

Результатом роботи мережі, в нашому випадку — останнього згорткового шару, буде карта ознак — 3-х вимірний тензор розмірністю 7 x 7 x 512. Перші дві розмірності є просторовими та відповідають за прояви тих чи інших ознак на сегменті зображення. Остання розмірність — це розмірність ознак, тобто мережа видала на вихід вектори розмірністю 512 для кожного з $7 * 7 = 49$ пікселів вхідного зображення. Просторові розмірності є непотрібними для подальшої роботи алгоритму, тому, щоб зменшити кількість обчислень, використовується Global Average Pooling — усереднення ознак за просторовими координатами карти ознак. Отже, результатом мережі буде вектор ознак розмірністю 512 для всього зображення.

АЛГОРИТМ ПЕРЕДБАЧЕННЯ РЕАКЦІЇ АУДИТОРІЇ НА ПОВІДОМЛЕННЯ

Розроблена нейронна мережа прогнозує порядок значення реакції користувачів на певне повідомлення, а не цю конкретну величину, оскільки ці значення мають значну дисперсію в лінійній шкалі і незначну в логарифмічній. Порядок величин, які нейронна мережа буде передбачати — це кількість лайків і кількість поширень. Таке обмеження дає можливість запропонованій мережі навча-

тися ефективніше. Однак прогнозування порядку, а не точної кількості реакцій не обмежує використання мережі з практичною метою, оскільки інформації про порядок кількості “лайків” та “поширень” достатньо для того, щоб прийти до рішення про публікацію чи відмову від публікації контенту.

Компоненти моделі. Зведемо регресійну задачу прогнозування лайків і репостів до задачі класифікації. Здійснюється це таким чином. Для кожної сторінки та автора є відомим розмір його аудиторії, який позначимо як N . Кількісні параметри i -ї публікації (P_i) — лайки та репости, позначимо як l_i та r_i відповідно. Для кожної публікації розрахуємо нормовані параметри відносно розміру аудиторії:

$$\forall P_i : l_i^* = \frac{l_i}{N}, r_i^* = \frac{r_i}{N}.$$

Маючи сукупність нормованих лайків для кожної публікації, рівномірно розіб'ємо їх на C_l інтервалів, довжина яких дорівнює $\frac{\max(l_i^*)}{C_l}$. Таким розбиттям ми фактично зводимо задачу регресії для лайків до задачі класифікації на C_l класів.

Аналогічно зводимо задачу регресії для поширень до задачі класифікації на C_r класів.

В наших експериментах використано такі значення: $C_l = 5$, $C_r = 5$.

Емоційна направленість коментарів, або сентимент-аналіз, також є задачею класифікації. Для тренування запропонованої моделі є необхідним здійснення розмітки коментарів за їхньою емоційною направленістю. Для збору цих даних використано вже відомий підхід, що базується на глибоких нейронних мережах. Дані опрацьовували таким чином. Кожен коментар публікації подавався на вхід мережі, яка видавала ймовірність p_j того, що коментар виражає позитивне ставлення користувача до контенту. Далі рахували середнє значення всіх ймовірностей p^* за всіма коментарями i , якщо $p^* > 0,5$, то загальна реакція аудиторії в коментарях на публікацію вважалась позитивною, інакше — негативною.

Отже, в рамках цієї роботи розв'язували три незалежні завдання класифікації відповідно до предмета аналізу — визначення та передбачення:

- 1) кількості лайків;
- 2) кількості поширень;
- 3) сентименту коментарів реакцій.

Дані трьох типів було надано на вхід алгоритму: зображення, додане в публікації; текст публікації; мета-інформація про сторінку та автора — математичне сподівання та дисперсія для лайків, поширень.

На вхідні дані накладено деякі обмеження: 1) обробляються публікації, що складаються з текстового опису і одного зображення, якщо зображень декілька — лише одне вибирається для аналізу; 2) аудіо та відео контенти ігноруються.

Ці обмеження не зменшують узагальненого характеру методу, а лише спрощують завдання і дають змогу простіше аналізувати поведінку моделі. Система може бути узагальнена для оброблення як відео-, так і аудіо-контенту.

Під час тренування моделі на дані накладалося додаткове обмеження — мережа тренувалася лише на тих публікаціях, для яких є не менше за 3 коментарі від аудиторії.

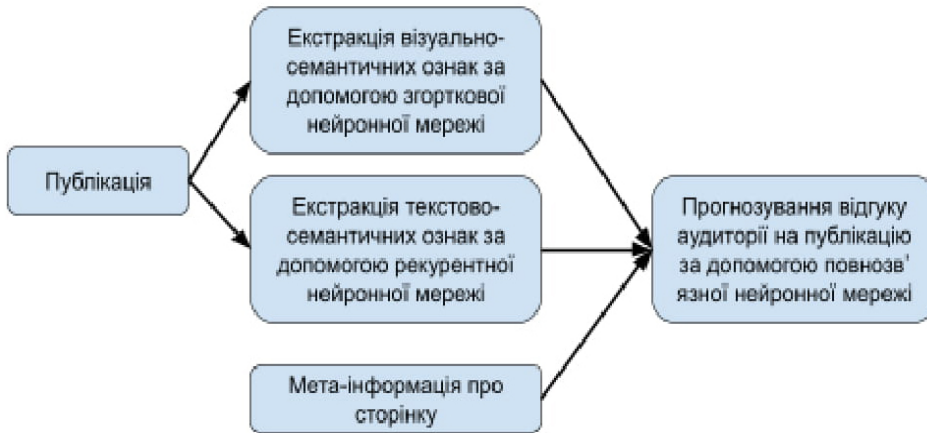


Рис. 4. Схема роботи алгоритму прогнозування реакції аудиторії

Розроблений алгоритм складається з функціонування трьох головних компонентів (Рис. 4):

- 1) згорткова нейронна мережа для екстракції візуально-семантичних ознак для зображення;
- 2) рекурентна нейронна мережа для екстракції текстово-семантичних ознак для текстових даних;
- 3) повнозв'язна нейронна мережа, що приймає мета-інформацію про сторінку та ознаки, отримані на попередніх кроках та генерує остаточне прогнозування.

Алгоритм працює таким чином.

1. Зображення нормалізується та подається на вхід згорткової нейронної мережі, результатом роботи якої буде згенерований вектор ознак розмірності 512.

2. Передоброблений текстовий вміст публікації подається на вхід рекурентної нейронної мережі, результатом роботи якої буде згенерований вектор ознак розмірності 768.

3. Отримані вектори та мета-інформація про сторінку об'єднуються в один вектор. Об'єднання візуально-текстових ознак та мета-інформації для прогнозування відбувається таким чином. Ознаки у вигляді векторів, отриманих на попередніх кроках, конкатенуються. Також до ознак додається ще 4 параметри — математичні сподівання та дисперсії кількості лайків та поширень, розраховані на основі попередніх публікацій автора. В результаті отримуємо фінальний вектор ознак розмірністю 1284.

4. Згенерований вектор ознак подається на наступний блок — нейронну мережу, що складається з трьох послідовних повнозв'язних шарів, основна мета яких — комбінувати візуальні, текстові ознаки та мета-інформацію про сторінку між собою і згенерувати фінальний вектор ознак. Далі три незалежні шари нейронної мережі на основі цих ознак прогнозують необхідні значення — порядок лайків і репостів та оцінку емоційної направленості коментарів аудиторії.

Для реалізації алгоритму використано мову програмування Python та бібліотеку PyTorch.

АПРОБАЦІЯ ПРОГНОЗУВАЛЬНОЇ МОДЕЛІ

Мережа тренувалася в end-to-end манері, за допомогою алгоритму оптимізації Adam. Оскільки задачу прогнозування було зведена до задачі класифікації трьох різних параметрів, використано категоріальну крос-ентропію як функція втрат:

$$CE = -\sum_{i=1}^C y_i \log(p_i),$$

де C — кількість класів; y_i — вектор, в i -й позиції якого буде 1, якщо об'єкт класифікації належить класу i , інакше 0; p_i — вектор, в i -й позиції якого буде ймовірність, що за якою об'єкт класифікації належить класу i .

Отже, якщо функцію втрат для прогнозування лайків позначити як L_{like} , репостів — L_{repost} , емоційної напруженості — $L_{\text{sentiment}}$, то кінцева функція втрат буде:

$$L_{\text{total}} = \alpha_1 L_{\text{like}} + \alpha_2 L_{\text{repost}} + \alpha_3 L_{\text{sentiment}},$$

де $\alpha_1, \alpha_2, \alpha_3$ — вагові коефіцієнти, що вказують на те, як сильно мережа має концентруватися на відповідному предметі прогнозування.

У експериментах з апробації прогнозувальної моделі значення параметрів були $\alpha_1 = 1, \alpha_2 = 0,5, \alpha_3 = 1$. Ці параметри вибрано такими, бо зазвичай прогнозувати поширення значно складніше у зв'язку з тим, що користувачі рідше поширюють публікації, ніж ставлять лайки і пишуть коментарі. Тому сильно штрафувати мережу за помилку, тобто збільшувати значення функції втрат, в цьому випадку є недоцільним.

Для аналізу процесу навчання мережі використовувався моніторинг зміни значень функції втрат в залежності від часу навчання, а саме — кількості епох навчання. Одна епоха — це епізод зміни ваг мережі, у якому кожен об'єкт з тренувального набору даних було використано один раз для коригування ваг мережі. Обчислення значення функції втрат на тренувальному наборі даних використовується для отримання значення помилки навчання та за використання валідаційного набору даних для отримання значення помилки узагальнення. У разі оптимального вибору моделі та алгоритму навчання, тренувальна і валідаційна помилки зменшуються за збільшення кількості епох навчання аж до порогового значення епохи, де вона максимально наближується до асимптотичного значення.

Під час тренування модель досягла мінімального значення на епосі навчання 17 і через 3 епохи навчання було зупинено, оскільки значення метрики функції втрат не покращувалось. Значення функції втрат на тренувальному та валідаційному наборах даних зі зростанням кількості епох навчання надано на рис. 5.

Значення функції втрат

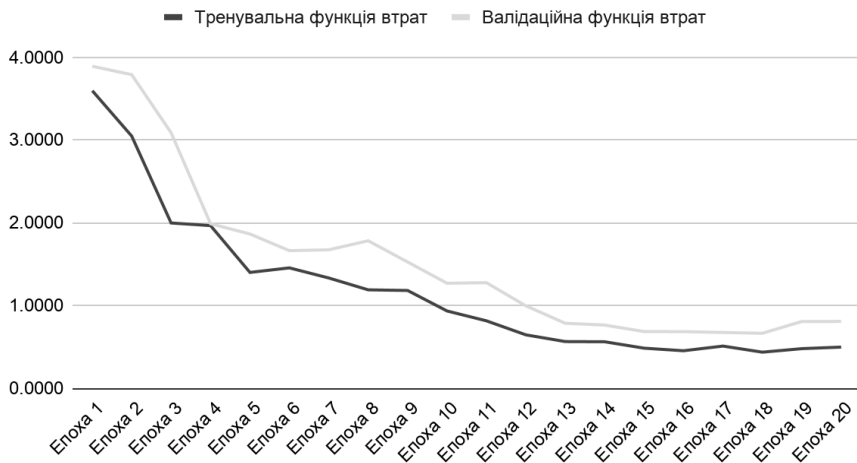


Рис. 5. Динаміка зміни значення функцій втрат (вісь y) в залежності від ітерацій навчання (вісь x)

Методика оцінювання якості алгоритму прогнозування реакції аудиторії. Для оцінювання якості роботи запропонованого алгоритму, а саме — якості прогнозування кількості лайків, поширень та сентимент-аналізу коментарів, використано такі метрики — точність (precision) та повнота (recall) [9]. Ці метрики використовуються у випадку, коли є дисбаланс у кількості даних різних класів. Метрики розраховуються для кожного класу, об'єкти, які належать до вибраного класу, називатимемо позитивами, а об'єкти інших класів — негативами. Для підрахунку метрик використано такі формули:

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN},$$

де TP — кількість об'єктів, які є позитивами, і модель класифікувала їх як позитиви, FP — кількість об'єктів, які є негативами, але модель їх класифікувала як позитиви; FN — кількість об'єктів, які є позитивами, але модель їх класифікувала як негативи.

Ці метрики показують якість роботи алгоритму з урахуванням помилок 1-го та 2-го роду (помилки типу “хибна тривога” та “пропуск цілі” відповідно).

Оцінювання якості роботи алгоритму здійснюється за такою методикою.

Крок 1 — встановлення мінімального порогу ймовірності віднесення об'єкту до певного класу. На основі розрахунків за моделлю для кожного підзавдання встановлюється такий мінімальний поріг ймовірності для певного класу, який має бути, щоб класифікувати досліджуваний об'єкт як об'єкт цього конкретного класу. Значення мінімального порогу лежить у діапазоні (0;1).

Крок 2 — побудова графіку залежності “точність-повнота” (precision-recall curve) за заданим мінімальним порогом. Варіюючи цей поріг, домагаємося зміни метрик кожного класу. Підвищення мінімального порогу зазвичай підвищує точність (precision), водночас повнота не зростає або спадає. Аналогічна ситуа-

ція виникає, якщо мінімальний поріг ймовірності знизити, тоді зростає повнота, а точність зменшується. Перебираючи значення мінімальних порогів для кожного класу, отримуємо графік залежності “точність-повнота” (precision-recall curve). Площа під кривою цього графіка — це число від 0 до 1; чим воно ближче до одиниці, тим якісніше працює алгоритм.

Крок 3 — визначення усередненої точності алгоритму для певного підзавдання (AP — average precision), яка розраховується як площа під кривою певного підзавдання.

Крок 4 — визначення загальної усередненої точності алгоритму (mAP - mean average precision), яка є усередненням за усіма завданнями кожного класу. Значення загальної усередненої точності алгоритму є показником якості роботи запропонованого алгоритму.

Аналіз результатів якості роботи. Кінцеві результати розрахунків за зазначеною методикою показано у таблиці 1.

Ми порівняли результати роботи розробленого нашого алгоритму з базовим алгоритмом, який видає випадковий результат на тестову публікацію.

За отриманими результатами визначено, що модель показала високу якість роботи під час прогнозування сентименту коментарів аудиторії. Якість прогнозування лайків і репостів є нижчою. Причини цього — великий рівень суб’єктивності, прояви можливої емоційності та упередженості аудиторії, небажання або відсутність можливості надати свою оцінку у вигляді лайка чи поширень, що збільшує дисперсію і не дає змогу точно спрогнозувати відповідне значення кількості лайків чи поширень.

Отже, побудовано нейронну мережу для прогнозування реакції аудиторії на контент, що створюють користувачі в соціальних мережах Фейсбук та Інстаграм. Для побудови і навчання нейронної мережі використано найточніші на сьогодні системи з автоматичного оброблення природної мови, а також оброблення візуальних даних (зображень). Кожна з цих систем була наперед навчена на великих обсягах даних і в цій роботі ми використали ці напрацювання для підвищення точності роботи запропонованої мережі.

Проведено аналіз контенту за набором даних і визначено необхідність окремого тренування нейронної мережі для кожної сторінки в соціальній мережі. Процес тренування оптимізовано таким чином, щоб тренування відбувалось тільки для набору зображень і текстів конкретно заданої сторінки.

Особливістю цієї розробки є те, що допомогою запропонованої нейронної мережі реалізовано можливість поєднаного дослідження візуально-го і текстового контентів для прогнозування реакції аудиторії.

Табл. 1. Якість роботи алгоритму на різних підзадачах

Задача	Загальна усереднена точність	
	Базовий алгоритм (%)	Запропонований алгоритм (%)
Прогнозування лайків	20	69,7
Прогнозування поширень	20	63,4
Сентимент-аналіз коментарів	50	91,2

ВИСНОВКИ

Використання розробленої нейронної мережі, яка складається з трьох пов'язаних шарів і на вхід отримує характеристики текстової та візуальної частин контенту та мета-дані про досліджувану сторінку, дає змогу оцінити реакцію користувачів на чергову публікацію конкретної сторінки в соціальній мережі. Характеристики текстової частини контенту попередньо формуються за допомогою нейронної мережі BERT, а характеристики візуальної частини контенту — за допомогою нейронної мережі VGG-16.

За аналізом результатів роботи нейронної мережі визначено такі обмеження поточної системи, зазначені у порядку зменшення їхньої значущості:

а) необхідність врахування загального інформаційного контексту щодо явища і/або предмету публікації в момент її створення;

б) необхідність врахування темпоральної інформації — порядку публікацій і часові проміжки між публікаціями.

ЛІТЕРАТУРА

1. De Fina A. Storytelling and audience reactions in social media. *Language in Society*. 2016. № 45. P. 473–498.
2. Gaspar R., Pedro C., Panagiotopoulos P., Seibt B. Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events. *Comput. Human Behav.* 2016. No.56. P. 179-191.
3. Cliche, M. (2017). BB_twtr at SemEval-2017 task 4: *Twitter sentiment analysis with CNNs and LSTMs*. Proceedings of the 11th international workshop on semantic evaluations (SemEval-2017). P. 573–580.
4. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., and Polosukhin I. *Attention is all you need*. In Advances in Neural Information Processing Systems. 2017. P. 6000–6010.
5. Devlin J., Chang M-W, Lee K., Toutanova K. *Bert: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint arXiv:1810.04805. 2018.
6. Simonyan K., Zisserman A., *Very Deep Convolutional Networks for Large-Scale Image Recognition*, arXiv e-prints. 2014.
7. Russakovsky O., *ImageNet Large Scale Visual Recognition Challenge*, arXiv e-prints, 2014.
8. He K., Zhang X., Ren S., Sun S. *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*. IEEE International Conference on Computer Vision (ICCV), 2015. P. 1026–1034.
9. Glorot X., Bengio Y. 2010. *Understanding the difficulty of training deep feedforward neural networks*. [Electronic resource] Journal of Machine Learning Research.
10. Bishop C. M. *Neural networks and machine learning*. Berlin: Springer, 1998. 353 p. (Nato ASI Subseries F).
11. He K., Zhang X., Ren S., Sun S. *Deep Residual Learning for Image Recognition*. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV:IEEE. 2016. P. 770–778.

Отримано 30.11.2020

Anisimov A.V., DSc (Phys & Math), Corresponding member
of National Academy of Sciences of Ukraine,
Dean of the Faculty of Computer Science and Cybernetics
ORCID: 0000-0002-1467-2006

e-mail: anatoly.v.anisimov@gmail.com

Bevza M.V., PhD student

ORCID: 0000-0002-2697-4968

e-mail: maksymbevza@gmail.com

Bobyl B.V., PhD student

ORCID: 0000-0002-9612-1071

e-mail: bobylbohdan@gmail.com

Taras Shevchenko National University of Kyiv

60, Volodymyrska st., Kyiv, 01033, Ukraine

PREDICTION OF AUDIENCE REACTION ON TEXT-VISUAL CONTENT USING NEURAL NETWORKS

Introduction. Social networks provide an ability to receive highly personalized experiences for their users, giving them an opportunity to follow pages of other users that publicize relevant and interesting content for them. Feedback on visual and textual content of the message is being received from followers by shares, likes and commentary. Prediction of the audience reaction on a particular message in social media becomes more and more of an important task, because the message is irreversible and there's a need to predict audience reactions before its publication.

The purpose of the paper is to build a system that can predict the reaction of the audience on the post and account for all the specialties of the page itself, its audience, the author and variety of possible reactions.

Results. We analyzed structure of the content and reactions in the social media and defined specialties of the messages, that should be accounted for during the analysis of the message popularity and reactions of the audience on it. We explain the process of the neural network training, that gives the ability to train the neural network for each particular page and audience to get better quality of the algorithms work. Features of the textual part of the content were obtained using the BERT neural network, whereas features of the visual part were obtained with the VGG-16 neural network. We've created the model that performs classification for three types of values: likes, shares and emotional focus of the commentaries. We have created a system that processes both visual-semantic and text-semantic parts as well as meta-information of the content and gives the program the full context of the publication that algorithm will process. Analysis of the developed system shows that model performs on emotional focus prediction better than amount of likes and shares.

Conclusions. The developed system uses three neural network layers and features from textual and visual parts of the content as well as meta-data of the message provides ability to predict reaction of the users on a particular message in the social media.

Keywords: artificial intelligence, natural language processing, computer vision, social networks, audience reaction prediction.