КІБЕРНЕТИКА

УДК 004.8

А.В. АНІСІМОВ
Київський національний університет імені Тараса Шевченка, Київ, Україна,
e-mail: *avatatan@gmail.com.*

О.О. МАРЧЕНКО
Київський національний університет імені Тараса Шевченка, Київ, Україна,
e-mail: *omarchenko@univ.kiev.ua.*

Е.М. НАСІРОВ
Міжнародний науково-навчальний центр інформаційних технологій та систем НАН та МОН України, Київ, Україна,
e-mail: enasirov@gmail.com.

В.Ю. ТАРАНУХА
Міжнародний науково-навчальний центр інформаційних технологій та систем НАН та МОН України, Київ, Україна,
e-mail: *taranukha@ukr.net.*

ПОРІВНЯЛЬНИЙ АНАЛІЗ НЕЙРОМЕРЕЖЕВИХ МОДЕЛЕЙ ДЛЯ ЗАДАЧ КЛАСИФІКАЦІЇ ТЕКСТІВ[1]

Анотація. У роботі досліджено задачу класифікації фейкових повідомлень та підходи до її розв'язання. Проведено порівняльний аналіз ефективності використання різних нейромережевих моделей для задач пошуку та класифікації фрагментів тексту, що містять неправдиві повідомлення. Досліджено вплив розмірності моделей на швидкість навчання, точність визначення та здатність адаптуватися до невідомих даних.

Ключові слова: штучний інтелект, комп'ютерна лінгвістика, нейромережа.

ВСТУП

У сучасну цифрову епоху проблема фейкових новин різко загострилася, що великою мірою спричинено поширенням соціальних мереж та месенджерів як основних каналів розповсюдження інформації. Ці платформи не лише надають користувачам безпрецедентні можливості для самовираження, але й забезпечують миттєвий доступ до великої кількості джерел часто сумнівної якості. Проблему поглиблює навмисне використання клікбейтів та інших методів, що коріняться в соціальній інженерії.

Розвиток соціальних мереж як джерела новин призвів до значного відходу від традиційних медіа, зокрема газет і телебачення. Звісно, традиційні джерела історично також дещо викривлювали подану інформацію, але вони мали й мають репутацію, яку потрібно підтримувати. Медіа з надійною репутацією намагалися донести правдиву інформацію або, принаймні, надавали її так, що часто було очевидно, хто і навіщо подає факти з певного погляду. Витрати на друк і поширення інформації у таких медіа досить високі, тому зазвичай вигідніше

[1] Дослідження виконано за грантової підтримки Національного фонду досліджень України в межах реалізації проєкту «Інформаційна технологія визначення тональності та класифікації текстової інформації для виявлення інформаційних загроз» (реєстраційний номер 2023.04/0053), представленого на конкурс «Наука для зміцнення обороноздатності України».

дотримуватися принципу надання достовірної інформації. У соціальних мережах ціна публікації мізерна, а ціна зіпсованої репутації також мінімальна. Завжди можна створити новий (напіванонімний) обліковий запис, особливо якщо джерело заявляє про певні його заборони чи утиски.

Явища фейкових новин класифікують у різні способи:

— класифікація за загальновизнаними типами дезінформації: клікбейт, часткове або повне введення в оману (фейки, включно з новинами), гумор і чутки [1–5]. Повідомлення всіх цих типів можуть спричинити лавинне (вірусне) поширення дезінформації. Проте вони дуже відрізняються за реалізацією та походженням, і тому досі немає надійного інструментарію, який ефективно розв'язує проблему у цьому контексті;

— класифікація за наміром і засобами: неточне інформування, дезінформація та розголошення конфіденційної інформації для завдання матеріальної, економічної, репутаційної шкоди, здійснення психологічного тиску тощо. Розголошення конфіденційної інформації для заподіяння шкоди по суті не є фейковими повідомленнями. Неточне інформування — апріорі неправдиве, а дезінформація — це якась неправдива інформація, надана з наміром зашкодити. Обидві останні категорії є фейковими повідомленнями [6]. Однак і цей підхід не вносить ясності у проблематику.

ПІДХОДИ ДО ВИЯВЛЕННЯ ФЕЙКОВИХ ПОВІДОМЛЕНЬ

Підходи до виявлення фейкових повідомлень також відрізняються за ефективністю та рівнем залучення людини у процес аналізу тексту.

Ручна перевірка. Це найстаріший, найповільніший, але також найнадійніший метод за умови, що його виконує кваліфікована досвідчена особа. На цьому ґрунтується Open Source Intelligence (OSINT) та її методи. Нині OSINT і фактчекери дуже поширені, аж до спеціалізованих телеграм-каналів [7]. Перевагою методу є те, що фахівець-людина враховує майже повний набір ознак, від фотографій і джерел до стилю та емоційного впливу. Недоліком методу є дуже низька швидкість процесу, тому його застосування досить обмежене.

Автоматична перевірка фактів на основі знань. Цей метод працює на основі встановлених баз даних фактів. Для цього потрібно створити та підтримувати базу даних триплетів «суб'єкт, предикат, об'єкт», а потім використовувати певний засіб побудови машинного виведення, щоб перевіряти, чи суперечать дані з підозрілого джерела встановленій істині. У цьому підході є багато проблем, зокрема потреба в уникненні повторів [8] і непостійна істинність тверджень чи фактів у часі: певне твердження чи факт сьогодні можуть бути істинними, а завтра хибними і навпаки [9], відповідно треба постійно підтримувати актуальність бази даних. Перевагою цього підходу є те, що кожен компонент системи взаємодіє прозоро, і процес перевірки можна виконувати незалежно.

Виявлення неправдивих повідомлень на основі аналізу стилю тексту. Цей підхід відповідає різним рівням сприйняття тексту: лексиці [10], синтаксичній структурі [11], дискурсу [12] і семантичному рівню [13]. Можна дійти висновку, що вищі рівні включають в себе більш прості рівні ціною зниження частоти спостережуваності ознак. Хоча аналіз на рівні семантики схожий з автоматичною перевіркою фактів на основі знань, однак різниця в методах призводить до значної різниці в результатах. Перевірка на основі знань є більш точною, тоді як семантичний метод на основі стилю може краще обробляти текст із новою інформацією та інформацією, не пов'язаною зі вмістом бази.

Виявлення фейкових повідомлень на основі аналізу настроїв (sentiment analysis). Ця група підходів є підмножиною методів виявлення на основі стилів і спирається на конкретні елементи стилю та ознаки, які використовує автор повідомлень, коли емоційно описує події [14]. Це надає цьому методу певну перевагу над загальними методами стилістичного аналізу повідомлень.

Аналіз характеристик поширення. Ця група методів ґрунтується на аналізі графів взаємозв'язків [15] і є унікальною, оскільки, на відміну від стилю, настрою, фактів, засобів мультимедіа тощо, фахівцям потрібні заздалегідь розроблені інструменти для аналізу графів. Тому це єдиний підхід, за якого машина краще виконує задачу, ніж людина.

Розвиток нейронних мереж дав змогу вдосконалити старі підходи та об'єднати різні функції в більш уніфіковані методи.

Векторизація. Такі засоби, як Glove та Word2Vec, збагатили старі методи та використовуються в нових підходах [16, 17]. Водночас системи трансформерної архітектури, наприклад BERT [18], дали змогу виконати тонке налаштування наявних попередньо навчених мереж та їхніх побудованих векторів для отримання остаточного результату.

Згорткові нейронні мережі (CNN). Вони були застосовані до задачі виявлення фейкових повідомлень з достатньо високими показниками, як це продемонстровано в [19]. Головною перевагою CNN є їхня здатність розглядати як окремі слова, так і послідовності слів, не вимагаючи повної точної відповідності та працюючи з відносно невеликими локалізованими фрагментами тексту поступово з подальшим об'єднанням результату.

Рекурентні нейронні мережі (RNN). Вони дають змогу фіксувати деякі послідовності ознак у часі з урахуванням поточного стану, оскільки в RNN дані використовуються рекурсивно. Коли CNN вимагає близької відповідності шаблону, RNN може зафіксувати відкладену (поза межами сприйняття) ознаку, якщо це важливо для розпізнавання. У [20] для розпізнавання фейкових повідомлень використано двоспрямований класифікатор LSTM. Крім того, використання гібридних підходів може значно підвищити продуктивність: гібридні системи (наведені у [17], із SNN, LSTM і вбудованими компонентами) значно перевершують системи з єдиним підходом.

Ще один спосіб підвищити продуктивність — включити інші джерела даних, насамперед зображення. До цього кожен компонент таких систем працював здебільшого незалежно. За такого підходу інформація з різних модальностей вноситься в результат лише за допомогою операцій кодування та об'єднання, що ускладнює передачу нетипової, малоінформативної (але значущої) та низькочастотної інформації між блоками, що оперують різними типами даних. У [21] стверджується, що було отримано високу продуктивність на конкретному мультимодальному наборі даних [22] за рахунок передачі якомога більшої кількості ознак між модулями оброблення даних різного типу так, що виникла цілісна мережа, де ознаки, отримані з тексту, допомагають розпізнавати зображення і навпаки.

У підсумку можна зазначити, що справжнім фейковим повідомленням слід вважати не просто недостовірну інформацію про певну подію, якої не було і яка завтра може настати і відповідна новина стане правдивою, а деяке неправдиве повідомлення, написане та розміщене навмисно з конкретною ціллю — для завдання певної інформаційної шкоди. Тому гіпотетично тексти цього типу можуть мати певні стилістичні та емоційні ознаки. Отже, методи стилометрії та сентимент-аналізу в поєднанні із сучасними нейронними мережами являють собою достатній набір засобів для досліджень задачі.

ПОРІВНЯЛЬНИЙ АНАЛІЗ НЕЙРОМЕРЕЖЕВИХ МОДЕЛЕЙ

Ціллю цієї статті є проведення порівняльного аналізу ефективності використання різних нейромережевих моделей для задач пошуку та класифікації фрагментів тексту, що містять неправдиві повідомлення (так звані fake news). Як базові архітектури для випробувань вибрано послідовну нейронну мережу SNN, довгу короткочасну пам'ять LSTM і трансформерну мережу BERT.

Вибір першої архітектури зумовлений потребою в перевірці того, наскільки послідовний порядок оброблення текстової інформації є достатнім і ефективним для пошуку та класифікації фрагментів, що містять недостовірну інформацію.

Довга короткочасна пам'ять LSTM є таким типом рекурентних нейронних мереж, які містять спеціальні засоби для запам'ятовування релевантної нової вхідної інформації, видалення вже непотрібних даних, отриманих з минулих тактів роботи мережі, інтеграції актуальних старих та нових даних в єдиний поточний контекст, а також для формування поточного виходу мережі. Ці засоби дають можливість ефективно обробляти довгі послідовності, що забезпечує значну перевагу LSTM над простими рекурентними нейронними мережами RNN у разі застосування до низки задач.

Архітектура BERT є двоспрямованою кодувальною трансформерною мовною моделлю, що містить 12 шарів блоків-трансформерів з 12 двоспрямованими багатоголовими шарами уваги (multi-head attention layer) кожен. Ця модель має близько 100 млн параметрів. Її навчали з використанням корпусу текстів BookCorpus (800 млн слів) і відфільтрованої версії англійської вікіпедії (2500 млн слів). Це дало змогу сформувати потужну векторну модель, в якій кожному слову в тексті присвоюється вектор, що описує його значення. За допомогою цих векторів можна точно порівнювати значення слів та встановлювати їхню схожість і ступінь подібності, а також семантичну зв'язність. Зауважимо, що подібні векторні представлення слів та засоби вимірювання їхньої семантичної близькості та зв'язності пропонували й до появи трансформерних мовних моделей, наприклад, в [23–26], проте слід зазначити, що векторна модель BERT має низку переваг. На відміну від уже зазначених методів, які генерують для кожного слова статичний вектор, що враховує всі значення, в яких це слово вжито в навчальному текстовому корпусі, BERT формує вектори динамічно, будуючи вектор для конкретного слова в тексті та враховуючи локальний контекст слів його найближчого оточення. Водночас розв'язується проблема омонімії та інші складнощі з багатозначністю слів. Двоспрямовані багатоголові механізми уваги уможливлюють врахування різноманітних складних відношень і зв'язків між словами на лексичному, синтаксичному та семантичному рівнях мови. Саме це дає підстави розраховувати на суттєві результати у разі застосування трансформерних великих мовних моделей (Large Language Models, LLM) для задач пошуку та класифікації фрагментів текстів, що містять неправдиві повідомлення.

В експериментах здійснено порівняння нейронних мереж різних типів, SNN, LSTM та BERT, у разі застосування до задачі пошуку та класифікації неправдивих новин (fake news). Під час досліджень вимірювали точність класифікації та швидкість навчання нейромережевих моделей.

Побудовано послідовну нейронну мережу SNN, що містить 8109 параметрів і складається з таких блоків:

• **Вкладений шар (embedding layer)**: 8000 параметрів. Кількість токенів у вхідній послідовності — 512, розмірності векторів вкладення — 16.

• **Глобальний середній пулінг (global average pooling1d layer)**: розмірність глобально усередненого вектора — 16.

• **Повнозв'язний шар (dense layer)**: 102 параметри. Кількість вихідних нейронів — 6.

• **Повнозв'язний шар (dense layer)**: 7 параметрів. Кількість вихідних нейронів — 1.

Для перевірки впливу підвищення розмірності моделі на точність збільшено кількість параметрів та додано ще один повнозв'язний шар. У такий спосіб отримано архітектуру, що містить 33113 параметрів і складається з таких блоків:

• **Вкладений шар (embedding layer)**: 32000 параметрів. Кількість токенів у вхідній послідовності — 512, розмірності векторів вкладення — 64.

• **Глобальний середній пулінг (global average pooling1d layer)**: розмірність глобально усередненого вектора — 16.

• **Повнозв'язний шар (dense layer)**: 1040 параметрів. Кількість вихідних нейронів — 16.

• **Повнозв'язний шар (dense layer)**: 68 параметрів. Кількість вихідних нейронів — 4.

• **Повнозв'язний шар (dense layer)**: 5 параметрів. Кількість вихідних нейронів — 1.

Для моделі LSTM також побудовано 2 архітектури аналогічної розмірності для коректності порівняння. Отримано архітектуру з 8861 параметром і такими шарами:

• **Вкладений шар (embedding layer)**: 8000 параметрів. Кількість токенів у вхідній послідовності — 512, розмірності векторів вкладення — 16.

• **Двоспрямований шар (bidirectional layer)**: 672 параметри. Кількість токенів — 512, розмірності вихідних векторів — 16.

• **Двоспрямований шар (bidirectional layer)**: 176 параметрів. Кількість вихідних нейронів — 4.

• **Повнозв'язний шар (dense layer)**: 10 параметрів. Кількість вихідних нейронів — 2.

• **Повнозв'язний шар (dense layer)**: 3 параметри. Кількість вихідних нейронів — 1.

Для посилення архітектури моделі LSTM збільшено кількість параметрів шарів та додано шар викидання (dropout layer), в якому випадковим чином вимикаються певні нейрони на кожному кроці навчання. У такий спосіб отримано модель з 34929 параметрами й такими шарами:

• **Вкладений шар (embedding layer)**: 16000 параметрів. Кількість токенів у вхідній послідовності — 512, розмірності вкладених векторів — 32.

• **Двоспрямований шар (bidirectional layer)**: 16640 параметрів. Кількість токенів — 512, розмірності вихідних векторів — 64.

• **Двоспрямований шар (bidirectional layer)**: 2208 параметрів. Кількість вихідних нейронів — 8.

• **Повнозв'язний шар (dense layer)**: 72 параметри. Кількість вихідних нейронів — 8.

• **Шар викидання (dropout layer)**: Кількість вихідних нейронів — 8.

• **Повнозв'язний шар (dense layer)**: 9 параметрів. Кількість вихідних нейронів — 1.

Хоча моделі LSTM традиційно демонструють кращі результати за більшої кількості параметрів порівняно з розглядуваною моделлю, проте були вибрані приблизно однакові розміри для забезпечення коректного порівняння.

Модель BERT — це попередньо навчена велика мовна модель, яку можна точно налаштувати (fine-tuning model) для різноманітних завдань NLP. Точне налаштування BERT — це процес адаптації попередньо навченої моделі BERT до конкретної задачі, наприклад, класифікації тексту, розпізнавання іменованих сутностей або генерації відповіді на запитання.

Як базову модель для налаштування вибрано bert-base-uncased model, що містить 12 прихованих шарів (блоків-трансформерів), 12 голів уваги (attention heads) у кожному шарі та близько 110 млн параметрів. Точне налаштування здійснено на двох останніх шарах мережі блоками розміром 32 тексти за крок.

Як цільову функцію у навчанні використано крос-ентропійну функцію втрат (cross-entropy loss function). Вона обчислює перехресну втрату ентропії між розрахованими прогнозами та цільовим значенням, що дає кращий результат для незбалансованого навчального набору:

$$L_{CE}(\hat{y}, y) = -\log p(y \mid x),$$

де y — клас вхідного повідомлення x.

Як оптимізаційний метод під час навчання застосовано AdamW (Adam with Weight Decay, Адам із регуляризацією ваг). Це вдосконалений варіант популярного оптимізатора Adam, який використовують для навчання нейронних мереж. Основна ідея AdamW полягає у тому, щоб правильно застосовувати регуляризацію ваг (weight decay), що дає змогу покращити узагальнювальну здатність моделі. При цьому використано такі налаштування оптимізатора:

1. lr: 5×10^{-5} — швидкість навчання (значення змін параметрів для кожного оновлення).

2. eps:10^{-8} — число епсилон алгоритму Адам для числової стабільності.

Модель BERT складається з блоку вкладення (embedding), блоку енкодера (Bert encoder), який містить 12 шарів, шару об'єднання (pooling layer), шару викидання (dropout layer) та лінійного шару (linear layer).

Щоб отримати базові показники для порівняння, використано наївний баєсівський класифікатор.

Для підвищення ефективності навчання та запобігання недонавчанню, а також для зменшення обсягу використовуваної пам'яті вибрано 500 слів, що трапляються найчастіше. Інші позначено як відсутні у словнику та замінено на спеціальний токен на позначення відсутності у словнику.

Для експериментів вибрано такі набори розмічених новин:

1. Набір WELFake Dataset (72134 новини, 51.44 % фейків), в якому об'єднано чотири популярні набори новин (Kaggle, McIntire, Reuters, BuzzFeed Political). Набір даних містить чотири стовпці: порядковий номер (починаючи з 0), заголовок (заголовок текстової новини), текст (зміст новини), та мітку (0 = фейк і 1 = справжня).

2. Набір News Articles Dataset з Kaggle (2097 новин, 38.19 % фейків). Цей набір даних містить новинні статті, пов'язані з бізнесом і спортом, з вебсайту https://www.thenews.com.pk з 2015 р. до сьогодні. Він включає заголовок конкретної статті, її зміст і дату.

Додатково виконано декілька кроків передоброблення текстів, а саме: видалення HTML-тегів, видалення пунктуації та спецсимволів, зведення до нижнього регістру, видалення стоп-слів, зведення слів до нормальної форми.

Для кожної вибраної моделі проведено дві серії експериментів:

1. Навчання та тестування на об'єднаному корпусі, що містить тексти з обох наборів новин.

2. Навчання моделей на текстах з одного набору новин і тестування точності визначення дезінформації на текстах з іншого набору.

Отже, побудовано такі набори текстів новин для навчання, валідації та тестування:

1. Дані з обох наборів додано в один об'єднаний корпус. Для навчання та порівняння результатів отриманий об'єднаний корпус поділено на три частини — для навчання, валідації та тестування. Це дає змогу уникнути перенавчання на рівні гіперпараметрів. Завдяки відокремленій частині «тестування» можна виконати перевірку на даних, які не використовувалися на етапі навчання моделі. Ця перевірка дає змогу переконатися в ефективності моделі на широкому спектрі даних, а не лише на тих, на яких проводять навчання та оптимізацію моделей.

2. У другому експерименті перший набір новин WELFake Dataset поділено на дві частини — для навчання та валідації. Для тестування використано другий набір новин News Articles dataset. Мета цього експерименту — перевірити ефективність досліджуваних моделей для практичного використання на даних з нових, не відомих моделі джерел.

АНАЛІЗ РЕЗУЛЬТАТІВ ЕКСПЕРИМЕНТІВ

В експериментах точність визначення класів повідомлень обчислено за формулою

$$Accuracy = \frac{Tp + Tn}{N},$$

де Tp — кількість правильно визначених неправдивих новин, Tn — кількість правильно визначених правдивих новин, N — кількість усіх новин у поточному корпусі текстів.

Для об'єднаного корпусу наївний баєсівський класифікатор показав точність 83 %, а для окремих наборів текстів навчання та тестування — лише 53 %.

Модель SNN продемонструвала високі результати. На об'єднаному корпусі її точність становить 91.26 % (табл. 1), якої достатньо для багатьох практичних завдань. Водночас вона швидко досягає свого максимуму ефективності й після декількох епох виходить на плато (рис. 1, 2).

Проте для окремих наборів текстів можна очікувано спостерігати значне зниження точності, яка становила лише 38.24 %.

Таблиця 1. Точність на тестовому наборі даних після навчання

Модель	Об'єднаний корпус текстів	Окремі набори текстів
Наївний баєсівський класифікатор	83.00 %	53.02 %
SNN (8000 параметрів)	91.26 %	38.24 %
SNN (32000 параметрів)	92.11 %	38.10 %
LSTM (8000 параметрів)	88.24 %	37.32 %
LSTM (32000 параметрів)	93.27 %	37.80 %
BERT	98.92 %	38.46 %

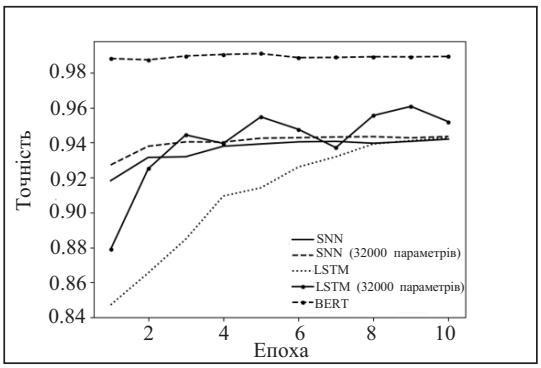

Рис. 1. Точність валідації моделей на різних епохах навчання

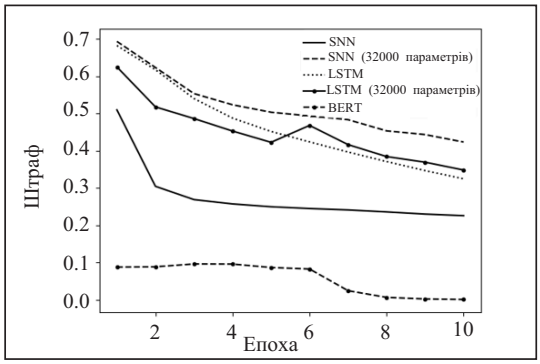

Рис. 2. Штраф моделей на різних епохах навчання

Збільшення кількості параметрів моделі SNN майже не вплинуло на результат, додавши до нього лише менше 1 %, відповідна точність становила 92.11 %. Водночас проблема швидкого насичення моделі досі є актуальною. Інакше кажучи, навіть менша версія моделі досягає максимуму точності, якого можна досягти за допомогою аналізу та знаходження послідовностей лінійних залежностей.

У разі окремих наборів текстів для більшої моделі отримано гірший результат, ніж для меншої. Точність моделі SNN з більшою кількістю параметрів становила 38.10 %.

Модель LSTM навчається повільніше, й лише з 8 епохи навчання виходить на плато, а також проявляються ознаки перенавчання — починаючи з певного моменту, точність на валідаційному наборі даних починає знижуватися. Точність на об'єднаному тестовому корпусі становила 88.24 %, а у разі окремих наборів текстів — 37.32 %. Це може свідчити про те, що використаної розмірності архітектури для більш складної моделі даних недостатньо для повного розкриття потенціалу. Для врахування контексту потрібно розглядати більший обсяг текстових особливостей, ніж той, що модель здатна запам'ятати.

У разі збільшення кількості параметрів моделі та додавання шару dropout для запобігання перенавчанню модель показала кращу точність 93.27 % на тестовому наборі даних з об'єднаного корпусу. У разі тестового набору з другого корпусу текстів її точність становила 37.80 %. Ці результати свідчать про здатність LSTM адаптуватися та враховувати контекст. Проте ця модель також має вищий штраф (рис. 2).

Загалом це може вказувати на проблеми, пов'язані з розміткою текстів або низькою репрезентативністю навчальних даних, але для перевіреного та збалансованого набору це, ймовірно, є ознакою тимчасового ефекту. У цьому разі модель може показати кращі результати під час подальшого навчання.

Таблиця 2. Час однієї епохи навчання досліджуваних моделей

Модель	Час однієї епохи навчання, с
SNN (8000 параметрів)	7.9
SNN (32000 параметрів)	11.3
LSTM (8000 параметрів)	465.3
LSTM (32000 параметрів)	673.1
BERT	3785.2

Базова модель BERT має найбільшу кількість шарів та параметрів з-поміж досліджуваних моделей. Через це її навчання потребує набагато більше часу, ніж навчання попередніх моделей (табл. 2). Проте результатом є отримана найкраща з-поміж досліджуваних моделей точність, яка становить 98.92 % на об'єднаному корпусі та 38.46 % на окремих текстових наборах. Водночас модель BERT демонструє високу точність з найменшим штрафом вже з перших епох навчання. Це може свідчити про здатність моделі враховувати найбільшу кількість наявних текстових ознак з-поміж усіх досліджуваних моделей.

Отже, модель SNN є достатньо ефективною для багатьох прикладних задач, забезпечує досить високу точність у разі швидкого навчання. Однак вона швидко досягає свого максимального потенціалу і виходить на плато.

Модель LSTM показує змішаний результат, оскільки менші моделі проявляють ознаки перенавчання. Для запобігання перенавчанню та підвищення точності потрібно збільшувати розмір моделі та використовувати шар викидання (dropout layer). Водночас ця модель потребує більше часу для навчання.

Модель BERT демонструє найвищі показники точності на об'єднаному корпусі, а також найкраще змогла пристосуватися до незнайомих даних. Проте вона потребує значно більше ресурсів та часу на навчання, ніж попередні моделі.

Значне зниження точності всіх досліджуваних моделей на різних наборах текстів для навчання та тестування свідчить про складність адаптування моделей до невідомих даних. Інакше кажучи, моделі, навчені на текстових даних, що містять певні особливості, притаманні одним авторам, новинним агенціям, групам проведення інформаційно-психологічних операцій, дають хибні результати, коли стикаються з текстами іншого авторства.

Варто зазначити, що нові достовірні новини та фейки створюються постійно різними та новими авторами, мають різне стилістичне та емоційне забарвлення, а також тематичне та смислове значення. Як наслідок, для практичного застосування досліджуваних моделей та підтримання високої точності їхньої роботи потрібно постійно продовжувати їхнє навчання на нових даних, що з'являються, та охоплювати максимально можливий спектр різних джерел та авторів.

ВИСНОВКИ

Проведені експерименти показали, що ефективність моделей може залежати від розміру та складності набору даних, а також від загальної репрезентативності навчальних даних. Використання глибокого навчання, зокрема моделей LSTM та BERT, може бути ефективним для аналізу текстових даних, оскільки вони здатні враховувати контекст і семантичні залежності. Проте можуть спостерігатися такі тимчасові негативні ефекти, як перенасичення моделі або зміна результатів у разі використання різних оброблюваних даних, і їх слід брати до уваги.

В експериментах, проведених для задачі класифікації недостовірних новин, найкращий стабільний результат показала донавчена модель BERT, точність якої становила 98.92 %. Водночас ця модель потребує найбільших обчислювальних ресурсів. У разі їхньої обмеженості як альтернативу можна використати LSTM-модель, яка також показує високу точність 96 %.

Продовження досліджень та розроблень у цій галузі уможливить подальше вдосконалення моделей та підвищення їхньої ефективності у разі застосування до реальних прикладних задач. Завдяки порівнянню та інтеграції різних підходів і методик можна буде виявити найбільш ефективні та надійні рішення для класифікації текстів у різних задачах як теоретичного, так і прикладного характеру.

СПИСОК ЛІТЕРАТУРИ

1. Daoud D.M., Abou El-Seoud S. An effective approach for clickbait detection based on supervised machine learning technique. *International Journal of Online and Biomedical Engineering.* 2019. Vol. 15, N 3. P. 21–32. https://doi.org/10.3991/ijoe.v15i03.9843.
2. Tacchini E., Ballarin G., Della Vedova M.L., Moret S., de Alfaro L. Some like it hoax: Automated fake news detection in social networks. arXiv:1704.07506v1 [cs.LG] 25 Apr 2017. https://doi.org/10.48550/arXiv.1704.07506.
3. Figueira A., Oliveira L. The current state of fake news: challenges and opportunities. *Procedia Computer Science.* 2017. Vol. 121. P. 817–825. https://doi.org/10.1016/j.procs.2017.11.106.
4. The Onion. Satirical newspaper. *Political satire since.* 1988. URL: https://www.theonion.com/.
5. Alkhodair S.A., Ding S.H., Fung B.C., Liu J. Detecting breaking news rumors of emerging topics in social media. *Information Processing & Management.* 2020. Vol. 57, Iss. 2. Article number 102018. https://doi.org/10.1016/j.ipm.2019.02.016.
6. Wardle C., Derakhshan H. Information disorder: toward an interdisciplinary framework for research and policy making. *Council of Europe report.* 2017. Vol. 27. P. 1–107. URL: https://rm.coe.int/information-disorder-report-november-2017/1680764666.
7. По той бік новин. URL: https://t.me/behindtheukrainiannews.
8. Maximilian N., Volker T., Hans-Peter K. Factorizing YAGO: scalable machine learning for linked data. *Proc. 21st international conference on World Wide Web.* (16–20 April 2012, Lyon, France). Lyon, 2012. P. 271–280. https://doi.org/10.1145/2187836.2187874.
9. Hoffart J., Suchanek F.M., Berberich K., Weikum G. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence.* 2013. Vol. 194. P. 28–61. https://doi.org/10.1016/j.artint.2012.06.001.
10. Mertoğlu U., Genc B. Lexicon generation for detecting fake news. arXiv:2010.11089v1 [cs.CL] 16 Oct 2020. https://doi.org/10.48550/arXiv.2010.11089.
11. Rubin V.L., Conroy N., Chen Y., Cornwell S. Fake news or truth? using satirical cues to detect potentially misleading news. *Proc. Second Workshop on Computational Approaches to Deception Detection* (17 June 2016, San Diego, California, USA). San Diego, 2016. P. 7–17. https://doi.org/10.18653/v1/W16-0802.
12. Karimi H., Tang. J. Learning hierarchical discourse-level structure for fake news detection. arXiv:1903.07389v6 [cs.CL] 10 Apr 2019. https://doi.org/10.48550/arXiv.1903.07389.
13. Bharadwaj P., Shao Z. Fake news detection with semantic features and text mining. *International Journal on Natural Language Computing (IJNLC)*. 2019. Vol. 8, N 3. P. 17–22. https://doi.org/10.5121/ijnlc.2019.8302.
14. Alonso M.A., Vilares D., Gomez-Rodriguez C., Vilares J. Sentiment analysis for fake news detection. *Electronics.* 2021. Vol. 10, Iss. 11. Article number 1348. https://doi.org/10.3390/electronics10111348.
15. Horne B.D., Nørregaard J., Adali S. Different spirals of sameness: A study of content sharing in mainstream and alternative media. *Proc. International AAAI Conference on Web and Social Media* (11–14 June 2019, Münich, Germany). Münich, 2019. Vol. 13 (01). P. 257–266. https://doi.org/10.1609/icwsm.v13i01.3227.
16. Gautam A., Jerripothula K.R. SGG: Spinbot, Grammarly and GloVe based fake news detection. *Proc. 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)* (24–26 September 2020, New Delhi, India). New Delhi, 2020. P. 174–182. https://doi.org/10.1109/BigMM50055.2020.00s033.

17. Ouassil M.A., Cherradi B., Hamida S., Errami M., el Gannour O., Raihani A. A fake news detection system based on combination of word embedded techniques and hybrid deep learning model. *International Journal of Advanced Computer Science and Applications.* 2022. Vol. 13, Iss.10. P. 525–534. https://doi.org/10.14569/IJACSA.2022.0131061.

18. Kaliyar R.K., Goswami A., Narang P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia Tools and Applications.* 2021. Vol. 80, Iss. 8. P. 11765–11788. https://doi.org/10.1007/s11042-020-10183-2.

19. Kaliyar R., Goswami A., Narang P., Sinha S. FNDNet— A deep convolutional neural network for fake news detection. *Cognitive Systems Research.* 2020. Vol. 61. P. 32–44. https://doi.org/10.1016/j.cogsys.2019.12.005.

20. Bahad P., Saxena P., Kamal R. Fake news detection using bi-directional LSTM-recurrent neural network. *Procedia Computer Science.* 2019. Vol. 165. P. 74–82. https://doi.org/10.1016/j.procs.2020.01.072.

21. Jing J., Wu H., Sun J., Fang X., Zhang H. Multimodal fake news detection via progressive fusion networks. *Information processing & management.* 2023. Vol. 60, Iss. 1. https://doi.org/10.1016/j.ipm.2022.103120.

22. Boididou C., Papadopoulos S., Kompatsiaris Y., Schifferes S., Newman N. Challenges of computational verification in social multimedia. *Proc. 23rd International Conference on World Wide Web* (7–11 April 2014, Seoul, Korea). Seoul, 2014. P. 743–748. https://doi.org/10.1145/2567948.2579323.

23. Mikolov T., Chen K., Corrado G., Dean J. Efficient estimation of word representations in vector space. arXiv:1301.3781v3 [cs.CL] 7 Sep 2013. https://doi.org/10.48550/arXiv.1301.3781.

24. Anisimov A.V., Marchenko O.O., Kysenko V.K. A method for the computation of the semantic similarity and relatedness between natural language words. *Cybernetics and Systems Analysis.* 2011. Vol. 47, N 4. P. 515–522. https://doi.org/10.1007/s10559-011-9334-2.

25. Marchenko O.O. A method for automatic construction of ontological knowledge bases. I. Development of a semantic-syntactic model of natural language. *Cybernetics and Systems Analysis.* 2016. Vol. 52, N 1. P. 20–29. https://doi.org/10.1007/s10559-016-9795-4.

26. Anisimov A.V., Marchenko O.O., Vozniuk T.G. Determining semantic valences of ontology concepts by means of nonnegative factorization of tensors of large text corpora. *Cybernetics and Systems Analysis.* 2014. Vol. 50, N 3. P. 327–337. https://doi.org/10.1007/s10559-014-9621-9.

A.V. Anisimov, O.O. Marchenko, E.M. Nasirov, V.Y. Taranukha

COMPARATIVE ANALYSIS OF NEURAL MODELS FOR TEXT CLASSIFICATION PROBLEMS

Abstract. The paper investigates the phenomenon of fake messages and approaches to their detection. A comparative analysis of the effectiveness of using different neural network models for the tasks of searching and classifying text fragments containing false messages is conducted. The influence of model's dimension on the learning speed, detection accuracy, and ability to adapt to unknown data is investigated.

Keywords: artificial intelligence, computational linguistics, neural network.

Надійшла до редакції 17.12.2024