

УДК 004.8

DOI <https://doi.org/10.32782/IT/2024-3-7>

Олег КОБИЛІН

кандидат технічних наук, доцент, завідувач кафедри інформатики, Харківський національний університет радіоелектроніки, пр. Науки, 14, м. Харків, Україна, 61166

ORCID: 0000-0003-0834-0475

Ірина ВЕЧІРСЬКА

кандидат технічних наук, доцент кафедри інформатики, Харківський національний університет радіоелектроніки, пр. Науки, 14, м. Харків, Україна, 61166

ORCID: 0000-0001-7964-2361

Анатолій АФАНАСЬЄВ

аспірант кафедри інформатики, Харківський національний університет радіоелектроніки, пр. Науки, 14, м. Харків, Україна, 61166

ORCID: 0009-0005-0707-981X

Бібліографічний опис статті: Кобилін, О., Вечірська, І., Афанасьєв, А. (2024). Аналіз існуючих моделей глибинного навчання в задачах обробки природної мови. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 3, 63–76, doi: <https://doi.org/10.32782/IT/2024-3-7>

АНАЛІЗ ІСНУЮЧИХ МОДЕЛЕЙ ГЛИБИННОГО НАВЧАННЯ В ЗАДАЧАХ ОБРОБКИ ПРИРОДНОЇ МОВИ

Обробка природної мови (NLP) є однією з найактуальніших галузей штучного інтелекту, що охоплює широкий спектр завдань, таких як аналіз емоцій, машинний переклад, розпізнавання мовлення та інші.

Мета роботи: Метою цього дослідження є всебічний аналіз продуктивності моделей глибинного навчання, включаючи рекурентні нейронні мережі (RNN), мережі довготривалої короткочасної пам'яті (LSTM) та керовані рекурентні блоки (GRU), у задачах NLP. Особлива увага приділяється ефективності цих моделей у завданнях аналізу емоцій.

Методологія: Дослідження включає кілька етапів: збір та попередню обробку даних, реалізацію та навчання моделей RNN, LSTM і GRU на вибраних наборах даних, оцінку їхньої продуктивності за допомогою таких показників, як точність, пригадування та F1-score, а також аналіз ресурсних вимог моделей, особливо в умовах обмежених обчислювальних ресурсів. Крім того, у роботі проводиться порівняльний аналіз моделей за показниками їхньої масштабованості при роботі з великими обсягами даних.

Наукова новизна: Дане дослідження пропонує детальний порівняльний аналіз ефективності RNN, LSTM та GRU в різних задачах NLP, з акцентом на їхній здатності обробляти послідовні дані та враховувати довготривалу залежність. Проведений аналіз виявляє, яка з моделей є найбільш ефективною в конкретних умовах, залежно від доступних ресурсів і специфіки даних.

Висновки: В результаті дослідження було встановлено, що GRU показала найвищу продуктивність в аналізі емоцій, перевершуючи RNN і LSTM за точністю, пригадуванням і F1-score. LSTM виявилася оптимальною для роботи з великими обсягами даних, демонструючи високу ефективність і точність. RNN, хоча і забезпечує швидке навчання на невеликих наборах даних, поступається іншим моделям у точності, що робить її менш придатною для складних задач NLP. Отримані результати містять цінну інформацію для дослідників і практиків, які займаються застосуванням моделей глибинного навчання у задачах NLP.

Ключові слова: глибинне навчання, рекурентні нейронні мережі, LSTM, GRU, обробка природної мови, аналіз емоцій.

Oleg KOBYLIN

Ph.D., Associate Professor, Head of the Department of Informatics, Kharkiv National University of Radio Electronics, 14, Nauky Ave., Kharkiv, Ukraine, 61166, oleg.kobylin@nure.ua

ORCID: 0000-0003-0834-0475

Iryna VECHIRSKA

Ph.D., Associate Professor at the Department of Informatics, Kharkiv National University of Radio Electronics, 14, Nauky Ave., Kharkiv, Ukraine, 61166, iryna.vechirska@nure.ua

ORCID: 0000-0001-7964-2361

Anatolii AFANASIEV

Postgraduate Student at the Department of Informatics, Kharkiv National University of Radio Electronics, 14, Nauky Ave., Kharkiv, Ukraine, 61166, anatolii.afanasiev@nure.ua

ORCID: 0009-0005-0707-981X

To cite this article: Kobylin, O., Vechirska, I., Afanasiev, A. (2024). Analiz isnuichykh modelei hlybynnoho navchannia v zadachakh obrobky pryrodnoi movy [Analysis of existing deep learning models in natural language processing tasks]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 3, 63–76, doi: <https://doi.org/10.32782/IT/2024-3-7>

ANALYSIS OF EXISTING DEEP LEARNING MODELS IN NATURAL LANGUAGE PROCESSING TASKS

Natural language processing (NLP) is one of the most relevant branches of artificial intelligence, covering a wide range of tasks such as emotion analysis, machine translation, speech recognition and others.

Purpose of work: *The purpose of this study is to comprehensively analyze the performance of deep learning models, including recurrent neural networks (RNN), long-short-term memory (LSTM) networks, and guided recurrent units (GRU), in NLP tasks. Special attention is paid to the effectiveness of these models in emotion analysis tasks.*

Methodology: *The study includes several steps: data collection and pre-processing, implementation and training of RNN, LSTM and GRU models on selected data sets, evaluation of their performance using indicators such as precision, recall and F1-score, and analysis of the resource requirements of the models, especially in conditions of limited computing resources. In addition, the paper provides a comparative analysis of models based on their scalability when working with large volumes of data.*

Scientific novelty: *This study offers a detailed comparative analysis of the performance of RNNs, LSTMs, and GRUs in various NLP tasks, with an emphasis on their ability to process sequential data and account for long-term dependencies. The conducted analysis reveals which of the models is the most effective in specific conditions, depending on the available resources and the specifics of the data.*

Conclusions: *As a result of the study, it was found that GRU showed the highest performance in emotion analysis, outperforming RNN and LSTM in terms of precision, recall and F1-score. LSTM proved to be optimal for working with large volumes of data, demonstrating high efficiency and accuracy. RNN, although it provides fast training on small data sets, is inferior to other models in accuracy, which makes it less suitable for complex NLP tasks. The obtained results contain valuable information for researchers and practitioners who are engaged in the application of deep learning models in NLP tasks.*

Key words: *deep learning, recurrent neural networks, LSTM, GRU, natural language processing, sentiment analysis.*

Актуальність проблеми. Обробка природної мови (NLP) стала ключовою галуззю штучного інтелекту, яка кардинально змінила взаємодію комп'ютерів з людською мовою. Вибух нових застосувань, таких як текстові генератори та інтелектуальні чат-боти, демонструє значний вплив NLP на різні сфери. Значні успіхи в цій галузі стали можливими завдяки моделям глибокого навчання, які ефективно вирішують складні лінгвістичні завдання (DeepLearning.AI, н.д.). Це дослідження зосереджене на аналізі ключових моделей глибокого навчання, зокрема RNN, LSTM та GRU, які широко застосовуються в NLP.

Аналіз останніх досліджень і публікацій. Зростання значення NLP у повсякденному житті відображається в його використанні в таких галузях, як роздрібна торгівля, медицина та віртуальні асистенти. Програми, такі як чат-боти, інструменти перекладу та системи аналізу настроїв, демонструють універсальність NLP. Розмовні агенти на кшталт Alexa і Siri стають

більш досконалими, а моделі, як GPT-3, вражають здатністю генерувати зв'язний текст. Google використовує NLP для покращення пошукової видачі, а соціальні мережі – для виявлення та фільтрації мови ворожнечі (James et al., 2013).

Моделі глибокого навчання здобули популярність завдяки здатності розпізнавати складні закономірності в лінгвістичних даних (Afanasieva et al., 2019). Це дослідження аналізує ефективність моделей RNN, LSTM і GRU у вирішенні різних задач NLP, таких як аналіз емоцій, машинний переклад, розпізнавання іменованих сутностей, виявлення спаму, генерація тексту та інші.

Незважаючи на значні досягнення, NLP ще стикається з такими проблемами, як упередженість моделей, незв'язність відповідей та періодична нестабільність поведінки. Це дослідження прагне оцінити сильні та слабкі сторони моделей глибокого навчання в завданнях NLP, надаючи цінну інформацію для дослідників і практиків (DeepLearning.AI, н.д.).

Мета дослідження. Мета цього дослідження полягає у всебічному та систематичному аналізі моделей глибокого навчання в галузі обробки природної мови. Досліджуються такі моделі, як рекурентні нейронні мережі, мережі довготривалої короткочасної пам'яті та керовані рекурентні блоки. Основна мета полягає в оцінці продуктивності цих моделей у різних задачах NLP, виявленні їхніх сильних і слабких сторін, а також аналізі нюансів у їхній продуктивності.

Цілі дослідження включають оцінку ефективності кожної моделі в задачах NLP, зокрема в аналізі емоцій, а також вивчення вимог до обчислювальних ресурсів кожної моделі з акцентом на масштабованість та ефективність. Це дозволить отримати уявлення про можливість їх практичної реалізації та виявити потенційні обмеження (Golian et al., 2022).

Методологія дослідження складається зі збору та попередньої обробки даних, реалізації та навчання моделей, використання показників ефективності для аналізу отриманих результатів та формулювання висновків.

Виклад основного матеріалу дослідження.

Рекурентні нейронні мережі (RNN) здатні ефективно обробляти послідовності різної довжини та моделювати часові залежності завдяки спільним параметрам, що дозволяє зменшити розмір моделі. Однак вони мають недоліки, такі як інтенсивність обчислень, складнощі з навчанням через проблему зникаючого градієнта та труднощі з обробкою довгих послідовностей.

Мережі довготривалої короткочасної пам'яті (LSTM) – це тип архітектури рекурентної нейронної мережі, розроблений для подолання проблем вивчення довготривалих залежностей у послідовних даних. У сфері обробки природної мови LSTM довела свою ефективність для таких завдань, як аналіз настроїв, переклад мови та розпізнавання емоцій (Nata, 2020).

LSTM добре підходять для задач, таких як аналіз настроїв, переклад мови та розпізнавання емоцій. Мережі LSTM ефективно

моделюють довготривалі залежності, демонструють стійкість до зашумлених даних і мають високу гнучкість у застосуванні. Проте вони характеризуються обчислювальною складністю, схильністю до перенавчання, складністю налаштування гіперпараметрів і високими вимогами до обсягу даних.

Керований рекурентний блок (GRU). Керований рекурентний блок – це тип архітектури рекурентної нейронної мережі, призначений для послідовної обробки даних (Nata, 2020). GRU є спрощеною версією LSTM з меншою кількістю параметрів, що робить їх обчислювально ефективнішими та легшими у навчанні. GRU використовують вентильні механізми для вибіркового оновлення прихованого стану, що дозволяє їм моделювати залежності в послідовних даних, але вони можуть бути менш ефективними для дуже довготривалих залежностей.

Мережі GRU мають меншу обчислювальну складність, ефективно моделюють довготривалі залежності та швидше навчаються. Однак вони можуть бути менш ефективними для дуже довготривалих залежностей, схильні до перенавчання та потребують ретельного налаштування гіперпараметрів.

Отримані результати (експериментальне підтвердження). Для експериментів було обрано набір даних розпізнавання емоцій «Emotion Dataset for Emotion Recognition Tasks» з відкритих джерел (Pandey, н.д.). Цей набір даних містить англомовні повідомлення з Twitter та використовується для завдань розпізнавання емоцій. Він включає повідомлення, що виражають чотири різних емоцій: гнів, страх, радість, та сум. На таблиці 1 наведені атрибути набору даних.

На рисунку 1 зображена частина даних (всього 3142) набору даних «Emotion Dataset for Emotion Recognition Tasks».

Експеримент 1 – Аналіз емоцій. Мета експерименту: аналіз емоцій забезпечує можливість виявлення та розпізнавання емоцій, виражених у текстових повідомленнях, що має значення в ряді сфер.

Таблиця 1

Атрибути набору даних

Атрибут	Опис
text (текст)	Текстова ознака, що містить повідомлення з Twitter. Це основний текст, який містить інформацію для аналізу емоцій.
label (мітка)	Класифікаційна мітка, що вказує на емоційний стан повідомлення. Має чотири можливих значень: – 0: гнів; – 1: страх; – 2: радість; – 3: сум.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	text,label														
2	You must be knowing #blithe means (adj.) Happy, cheerful,.joy														
3	Old saying 'A #smile shared is one gained for another day' @YEGlifer @Scott_McKeen,joy														
4	Bridget Jones' Baby was bloody hilarious pu... #BridgetJonesBaby,joy														
5	@Elaminova sparkling water makes your life sparkly,joy														
6	I'm tired of everybody telling me to chill out and everything's ok. no the fuck its not. I'm tired of faking a a fucking smile,joy														
7	#GBBO can cheer me up senéU,joy														
8	& as much as I hate for a dude to cheat, women forego pleasing their man, they get lazy & men get lazy & then it's just pointless,joy														
9	@GOT7Official @jrjyp happy birthday jin young!!!!!! #PrinceJinyoungDay #happyjinyoungday #got7 #birthday,joy														
10	@GOT7Official @jrjyp happy birthday jin young!!!!!! #PrinceJinyoungDay #happyjinyoungday #got7 #happy #birthday,joy														
11	The race advances only by the extra achievements of the individual. You are the individual. ~Charles Towne\n #optimism #inspire,joy														
12	The race advances only by the extra achievements of the individual. You are the individual. ~Charles Towne\n #inspire,joy														
13	Watching football matches without commentary is something that I rejoice, found a transmission of CityvB™s match like that today, joyful,.joy														
14	#twtd comes on soon #happy,joy														
15	#twtd comes on soon ,joy														
16	@TauDeltaPhiDK THANK YOU FOR MY OBAMA CUT OUT!!!!!! I am elated that he's back homepyR,joy														
17	@DDogsScout 'Oh! Almost with odd cheerfulness, Big Boss offers: 'Muzzle flash blinding. Accidental by the guy who became my best friend.' joy														
18	Gemma Simmons is the bright spot of the premiere so far. #AgentsofSHIELD,joy														
19	This is a beautiful day that the Lord has made. I will rejoice and be glad!!!,joy														
20	Watch this amazing live.ly broadcast by @kelli.peterson #lively #musically. Come WATCH,joy														
21	Sometimes I like to talk about my sadness. Other times, I just want to be distracted by friends, laughter, shopping, eating... \n\n#MHChat,joy														
22	Oi @THEWIGGYMESS you've absolutely fucking killed me.. 30 mins later im still crying with laughter.. Grindah.. Grindah... pud" hahahahahahaha,joy														
23	So happy I live in NYC! See you tomorrow@SamHeughan @Barbour,joy														
24	Accept the challenges, so that you may feel the exhilaration of victory.,joy														
25	For what a beautiful day. #elated,joy														
26	For what a beautiful day. ,joy														
27	@Langston_Hunter Yeah bro It eas hella exhilarating,joy														
28	sometimes im sad then remember Margaret Thatcher is dead and then I rejoice,joy														
29	This is not me brown nosing but I've listened to lots of housing ministers but @GavinBarwellMP #nhf16 impressed me more than any ,joy														
30	This is not me brown nosing but I've listened to lots of housing ministers but @GavinBarwellMP #nhf16 impressed me more than any #optimism,joy														
31	@APkrawczynski Any possibly KG is being bought out as a player so that he can buy in from Glen as a minority owner? #optimism,joy														
32	Martin Ro A shoes the effects of #Brexit: bright and talented foreigners made to feel unwelcome and leaving U.K. #c4news,joy														
33	one thing I can say is that you kept me smiling,joy														
34	#blackish always has me #rollin #hilarious,joy														
35	#blackish always has me #rollin ,joy														
36	How is your toddler coping with the new arrival? Has he tried killing her yet?' they cheerfully ask.,joy														
37	American Schools are lively,joy														
38	Drawing n foldine mini-comics is meditative and relaxine. I think I need to do them for more than just Halloween....,joy														

Рис. 1. Фрагмент набору даних «Emotion Dataset for Emotion Recognition Tasks»

Методика експерименту: використання алгоритмів машинного навчання для побудови моделей, які можуть автоматично класифікувати текст за емоційними категоріями.

Метрики: асигурація (точність) визначає, наскільки точно модель класифікує емоції; precision вимірює, яку частку емоцій модель класифікує правильно відносно загальної кількості визначених моделлю емоцій; recall (пригадування) показує, яку частку емоцій модель виявляє відносно загальної кількості існуючих емоцій; F1-score (F1-оцінка) – комбінація точності та пригадування, що надає комплексну оцінку продуктивності моделі.

Експеримент з аналізу емоцій дозволяє створити та вдосконалити моделі для виявлення емоцій у тексті, що має широкий спектр застосувань від соціальних мереж до різних галузей досліджень (Afanasieva et al., 2023).

Для написання коду для експериментів було використано мову програмування Python, бібліотеки та Jupyter Notebook для візуалізації результатів.

Рекурентні нейронні мережі (RNN). На першому етапі експерименту було завантажено набір даних із файлу 'emotion-labels-test.csv' за допомогою бібліотеки Pandas. Вхідні дані було розділено на текстові описи (X) та відповідні мітки класів (y). Для перетворення

категоріальних емоційних міток у числовий формат було застосовано метод Label Encoding. Після цього дані було поділено на тренувальний і тестовий набори для забезпечення коректної оцінки продуктивності моделі.

Далі було здійснено токенизацію текстових даних та додано «padding», щоб усі послідовності мали однакову довжину. Для побудови моделі RNN було використано Embedding-шар, SimpleRNN-шар та Dense-шар. Як функцію втрат було обрано бінарну крос-ентропію, а для оцінки точності моделі – відповідну метрику.

Після тренування модель RNN використовується для прогнозування емоцій на тестовому наборі даних. Для оцінки ефективності моделі було розраховано такі показники, як точність (асигурація), точність класифікації (precision), пригадування (recall) та F1-оцінка (F1-score), що дало змогу здійснити комплексний аналіз продуктивності моделі задачі в аналізі емоцій.

Результати виконання експерименту (рис. 2-4) свідчать про те, що модель RNN неефективно справляється із завданням аналізу емоцій.

Проаналізуємо основні метрики. Точність моделі є надзвичайно низькою, становлячи приблизно 22%, що вказує на неефективність моделі у розрізненні різних категорій емоцій. Це може свідчити про упередженість моделі, яка

```

mienshykova_experiment1_RNN.ipynb X
mienshykova_experiment1_RNN.ipynb > import pandas as pd
+ Code + Markdown | Run All | Clear All Outputs | Outline ...

Epoch 1/5
79/79 [=====] - 2s 18ms/step - loss: nan - accuracy: 0.3191 - val_loss: nan - val_accuracy: 0.2210
Epoch 2/5
79/79 [=====] - 1s 8ms/step - loss: nan - accuracy: 0.2471 - val_loss: nan - val_accuracy: 0.2210
Epoch 3/5
79/79 [=====] - 1s 8ms/step - loss: nan - accuracy: 0.2471 - val_loss: nan - val_accuracy: 0.2210
Epoch 4/5
79/79 [=====] - 1s 8ms/step - loss: nan - accuracy: 0.2471 - val_loss: nan - val_accuracy: 0.2210
Epoch 5/5
79/79 [=====] - 1s 8ms/step - loss: nan - accuracy: 0.2471 - val_loss: nan - val_accuracy: 0.2210
20/20 [=====] - 0s 2ms/step
Accuracy: 0.22098569157392686
Precision: 0.04883467580040673
Recall: 0.22098569157392686
F1-score: 0.07999221648118708
Classification Report:
      precision    recall  f1-score   support

0         0.22      1.00      0.36       139
1         0.00      0.00      0.00       201
2         0.00      0.00      0.00       167
3         0.00      0.00      0.00       122

 accuracy          0.22      629
 macro avg         0.06      0.25      0.09      629
 weighted avg      0.05      0.22      0.08      629
    
```

Рис. 2. Результати продуктивності моделі RNN на тестовому наборі даних

```

Confusion Matrix:
[[139  0  0  0]
 [201  0  0  0]
 [167  0  0  0]
 [122  0  0  0]]
    
```

Рис. 3. Матриця плутанини для моделі RNN на тестовому наборі даних

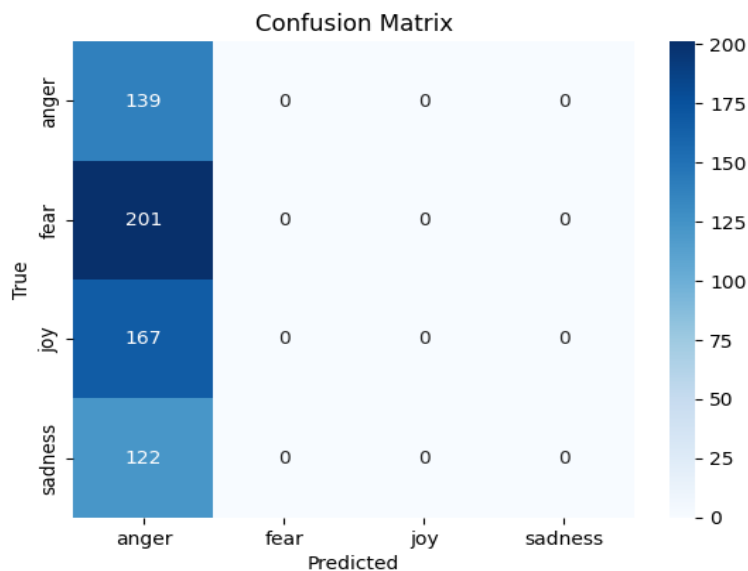


Рис. 4. Графік зміни втрат і точності моделі RNN під час навчання

схильна відносити більшість прикладів до певного класу.

Точність, пригадування та оцінка F1 для кожної категорії емоцій (0, 1, 2, 3) також виявилися дуже низькими, часто близькими до нуля. Це

свідчить про те, що модель не здатна коректно ідентифікувати екземпляри кожного класу емоцій, що підтверджується низьким показником пригадування, який вказує на значну кількість пропущених істинно позитивних прикладів.

Матриця плутанини підтверджує низьку продуктивність моделі, демонструючи, що більшість прикладів прогнозуються як такі, що належать до класу 0. Такий дисбаланс у прогнозах призводить до низьких показників точності, пригадування та F1-оцінки для інших класів.

Зокрема, низька точність вказує на те, що коли модель прогнозує екземпляр як позитивний, це часто виявляється невірним. Низьке пригадування свідчить про те, що модель пропускає значну кількість реальних позитивних прикладів.

Мережі довготривалої короткочасної пам'яті (LSTM). Процес реалізації для моделі LSTM був аналогічний до процесу для RNN. Після завантаження та підготовки даних було створено модель LSTM, яка включала Embedding-шар, LSTM-шар та Dense-шар для багатокласової класифікації. Як функцію втрат використовували категоріальну крос-ентропію, а точність слугувала основною метрикою для оцінки продуктивності. Модель LSTM була використана для отримання прогнозів на тестовому наборі даних.

Результати виконання експерименту (рис. 5-7) показали значне покращення

порівняно з попередньою моделлю, досягнувши точності приблизно 73,77%.

Проаналізуємо основні метрики. Модель продемонструвала загальну точність 73,77%, що вказує на її здатність коректно передбачати емоційні категорії.

Точність, пригадування та оцінка F1. Модель LSTM досягла високих показників точності, пригадування та F1-оцінки для класів 1 і 2, що свідчить про її ефективність у класифікації твітів, які належать до цих емоційних категорій. Однак клас 0 (емоційна категорія 0) демонструє нижчі показники точності, пригадування та F1-оцінки порівняно з іншими класами, що вказує на певні труднощі моделі у розпізнаванні цієї категорії. Клас 3 (емоційна категорія 3) показав помірні результати за всіма метриками.

Матриця плутанини. Матриця плутанини надає детальну розбивку прогнозів моделі для кожного класу, показуючи кількість істинно-позитивних, істинно-негативних, хибно-позитивних і хибно-негативних прогнозів. Клас 1 (емоційна категорія 1) мав найбільшу кількість істинно-позитивних прогнозів, що свідчить про ефективність моделі в ідентифікації цієї емоційної категорії. Водночас клас 0 (емоційна категорія 0)

```

mienshykova_experiment1_LSTM.ipynb X
mienshykova_experiment1_LSTM.ipynb > import pandas as pd
+ Code + Markdown | Run All | Restart | Clear All Outputs | Variables | Outline ...
... Epoch 2/5
79/79 [=====] - 1s 16ms/step - loss: 1.1991 - accuracy: 0.5388 - val_loss: 1.1421 - val_accuracy: 0.6216
Epoch 3/5
79/79 [=====] - 1s 17ms/step - loss: 1.0780 - accuracy: 0.7986 - val_loss: 0.9457 - val_accuracy: 0.7027
Epoch 4/5
79/79 [=====] - 1s 17ms/step - loss: 0.9664 - accuracy: 0.8456 - val_loss: 0.8382 - val_accuracy: 0.6630
Epoch 5/5
79/79 [=====] - 1s 17ms/step - loss: 0.5628 - accuracy: 0.8989 - val_loss: 0.7488 - val_accuracy: 0.7377
20/20 [=====] - 0s 5ms/step
Accuracy: 0.7376788553259142
Precision: 0.7594804234010711
Recall: 0.7376788553259142
F1-score: 0.7332648804205049
Classification Report:

```

	precision	recall	f1-score	support
0	0.88	0.55	0.68	139
1	0.66	0.90	0.76	201
2	0.83	0.79	0.81	167
3	0.69	0.61	0.65	122
accuracy			0.74	629
macro avg	0.77	0.71	0.72	629
weighted avg	0.76	0.74	0.73	629

Рис. 5. Результати продуктивності моделі LSTM на тестовому наборі даних

```

Confusion Matrix:
[[ 76  32  12  19]
 [  1 181  8  11]
 [  2  29 132  4]
 [  7  33  7  75]]

```

Рис. 6. Матриця плутанини для моделі LSTM на тестовому наборі даних

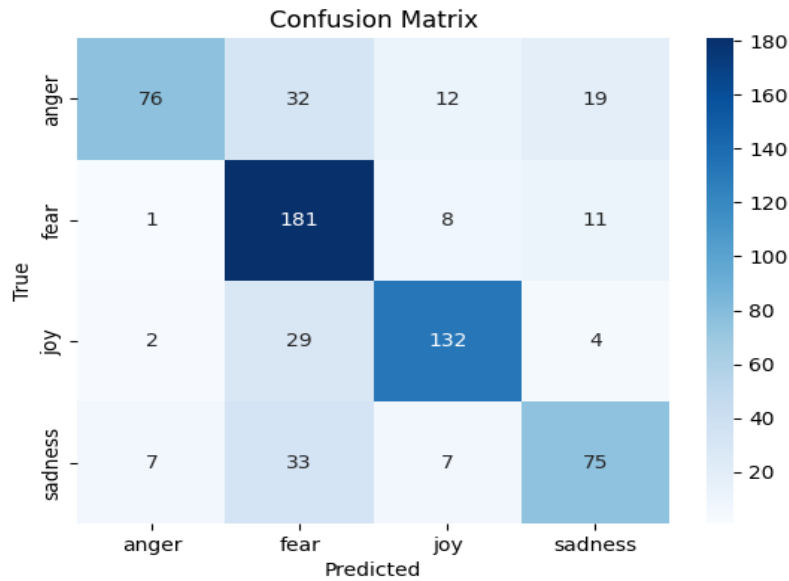


Рис. 7. Графік зміни втрат і точності моделі LSTM під час навчання

мав відносно високу кількість хибнонегативних результатів, що свідчить про тенденцію моделі пропускати випадки цієї емоційної категорії.

Керований рекурентний блок (GRU).

У цьому експерименті було використано той самий набір даних, що й у попередніх моделях. Спочатку визначено розмір словника (vocab_size), розмір вектора вбудовування (embedding_dim) та максимальну довжину послідовності (max_length). Текстові дані було токенізовано, а послідовності доповнено до максимальної довжини. Для побудови моделі GRU використано Embedding-шар, GRU-шар з поверненням послідовності, GlobalMaxPooling1D та Dense-шар для багатокласової класифікації. Як функцію втрат було обрано категоріальну крос-ентропію, а точність слугувала основною метрикою оцінки

моделі (Turuta et al., 2024). Модель навчалася на тренувальних даних протягом 10 епох з використанням валідаційного набору.

Результати виконання експерименту (рис. 8-10) показують, що модель GRU продемонструвала високу точність, пригадування та F1-оцінку.

Проаналізуємо основні метрики. Точність. Загальна точність моделі становить 82,22%, що свідчить про правильне передбачення емоційних категорій для 82,22% твітів у тестовому наборі. Середньозважена точність дорівнює 82,88%, а точність для окремих класів варіюється від 73% до 89%. Особливо високі показники спостерігаються для класів 0 і 2.

Пригадування. Середньозважений показник пригадування становить 82,22%, причому для

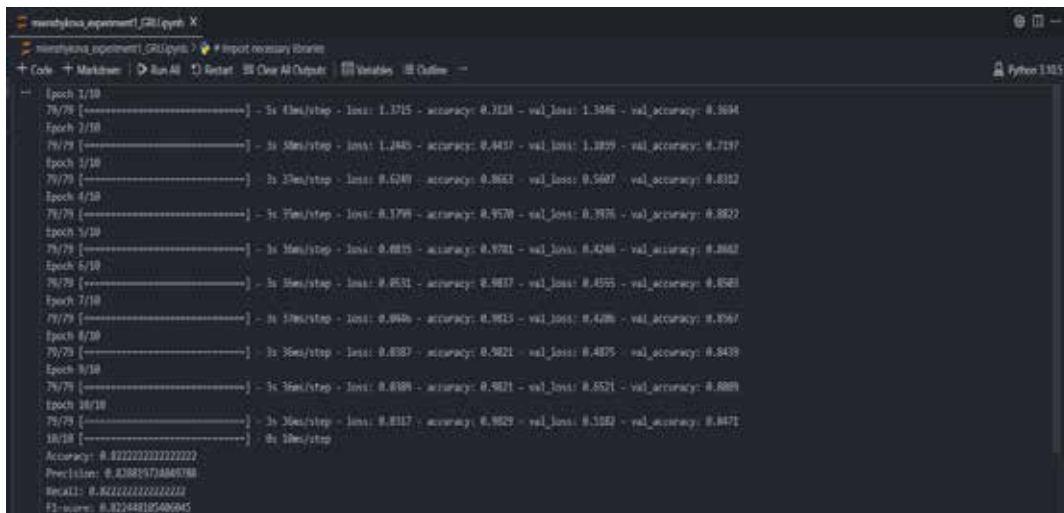


Рис. 8. Результати продуктивності моделі GRU на тестовому наборі даних

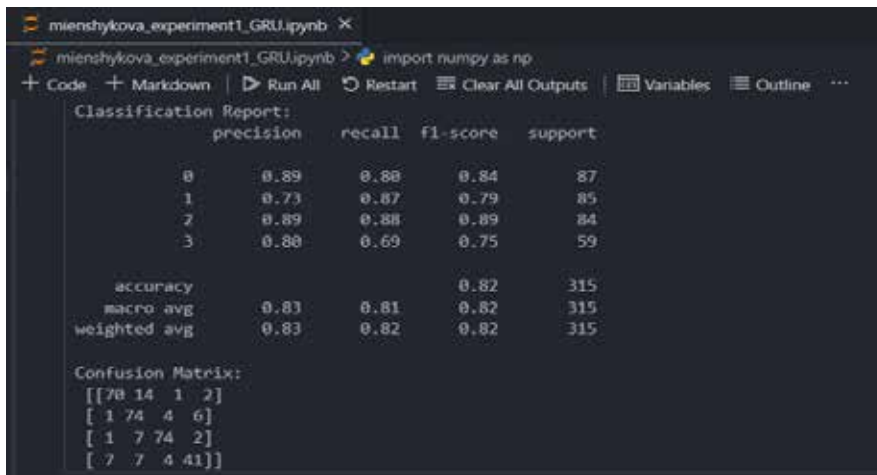


Рис. 9. Матриця плутанини для моделі GRU на тестовому наборі даних

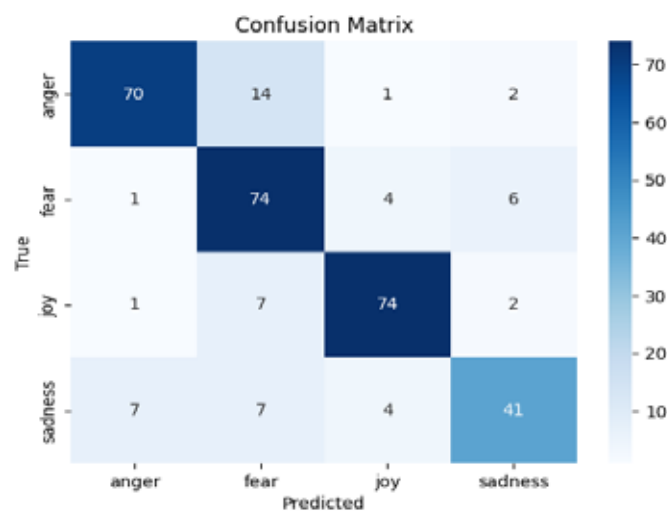


Рис. 10. Графік зміни втрат і точності моделі GRU під час навчання

окремих класів цей показник коливається від 69% до 88%. Найвищий рівень пригадування зафіксовано для класу 2.

Оцінка F1. Середньозважений показник F1 дорівнює 82,24%, що є гармонійним середнім між точністю та пригадуванням, і забезпечує збалансовану оцінку ефективності моделі.

Матриця плутанини. Матриця плутанини надає детальну розбивку прогнозів моделі для кожного класу. Вона показує, що модель працює добре для класів 0 і 2, проте має труднощі з класами 1 і 3, зокрема, плутає клас 1 з класом 3. Клас 2 (емоційна категорія 2) демонструє високу точність, пригадування та F1-оцінку. Водночас клас 1 (емоційна категорія 1) має нижчі показники у порівнянні з іншими класами, що вказує на потребу у вдосконаленні моделі для покращення класифікації цього класу. Матриця плутанини допомагає визначити конкретні області, де модель може потребувати

вдосконалення, наприклад, зменшення кількості помилкових класифікацій між певними класами (James et al., 2013).

Експеримент 2 – Тест масштабованості.

Тест масштабованості є важливим інструментом для оцінки ефективності моделі при збільшенні розміру даних. Він дозволяє зрозуміти, як змінюються час тренування та використання ресурсів, що особливо актуально у випадках, коли розмір набору даних може варіюватися.

Мета тесту масштабованості включає дослідження впливу збільшення розміру даних на час тренування моделі, що допомагає визначити, чи існує лінійна або нелінійна залежність між розміром даних та часом тренування, а також оцінку того, як збільшення розміру даних впливає на точність моделі, що дозволяє зрозуміти, чи поліпшується точність при збільшенні обсягу даних.

Рекурентні нейронні мережі. Результати виконання експерименту (рис. 11) показують, що із збільшенням розміру набору даних час навчання також зростає, що є очікуваною поведінкою. Це пояснюється тим, що більші набори даних вимагають більше обчислювальних ресурсів та часу для обробки.

Збільшення часу навчання не є лінійним і може залежати від таких факторів, як складність моделі та ефективність апаратного забезпечення. Точність моделі покращується зі збільшенням розміру набору даних, що свідчить про те, що більший обсяг даних сприяє кращій роботі моделі. Проте покращення точності не є строго лінійним, що вказує на можливе зменшення прибутковості від збільшення розміру набору даних. Це явище є поширеним у машинному навчанні, де початкові прирости точності більш значущі, ніж наступні.

Загалом, результати експерименту демонструють, що збільшення розміру набору даних позитивно впливає на точність моделі, однак для практичних застосувань важливо знайти баланс між розміром набору даних та наявними обчислювальними ресурсами.

Мережі довготривалої короткочасної пам'яті. LSTM показали очікувані результати в експерименті зі збільшення розміру набору даних (рис. 12). Час навчання зростає із збільшенням обсягу даних, що цілком прогнозовано, оскільки більші набори даних вимагають більше обчислювальних ресурсів. Збільшення часу навчання не було строго лінійним, але спостерігалось помітне зростання.

Що стосується точності, модель показала покращення результатів на тестовому наборі

даних зі збільшенням обсягу навчальних даних. Менші набори даних, такі як 100 або 500 зразків, демонстрували нижчу точність, що було очікувано через обмеженість даних для навчання. Значне покращення точності спостерігалось при використанні більших наборів даних, таких як 10 000 або 20 000 зразків. Це вказує на те, що для невеликих наборів даних модель може не мати достатньо інформації для ефективного узагальнення, що призводить до зниження точності. У той же час, зі збільшенням розміру набору даних модель отримує більше прикладів для навчання, що позитивно впливає на її продуктивність.

Раптове підвищення точності для великих наборів даних може свідчити про здатність моделі навчатися більш складним закономірностям зі збільшенням обсягу даних. Однак при виборі розміру набору даних слід враховувати доступні ресурси і необхідність знайти баланс між часом навчання і точністю моделі.

Загалом, результати свідчать про те, що в цьому експерименті більший розмір набору даних призводить до кращої продуктивності моделі LSTM, проте при прийнятті рішень щодо оптимального розміру набору даних необхідно враховувати співвідношення між часом навчання і точністю.

Керований рекурентний блок. GRU у цьому експерименті показав очікувані результати зі збільшенням розміру набору даних (рис. 13). Час навчання моделі зростає із збільшенням обсягу даних, що є прогнозованим, оскільки більші набори даних вимагають більше обчислень для тренування моделі. На часову складність навчання впливає кількість зразків у наборі даних.

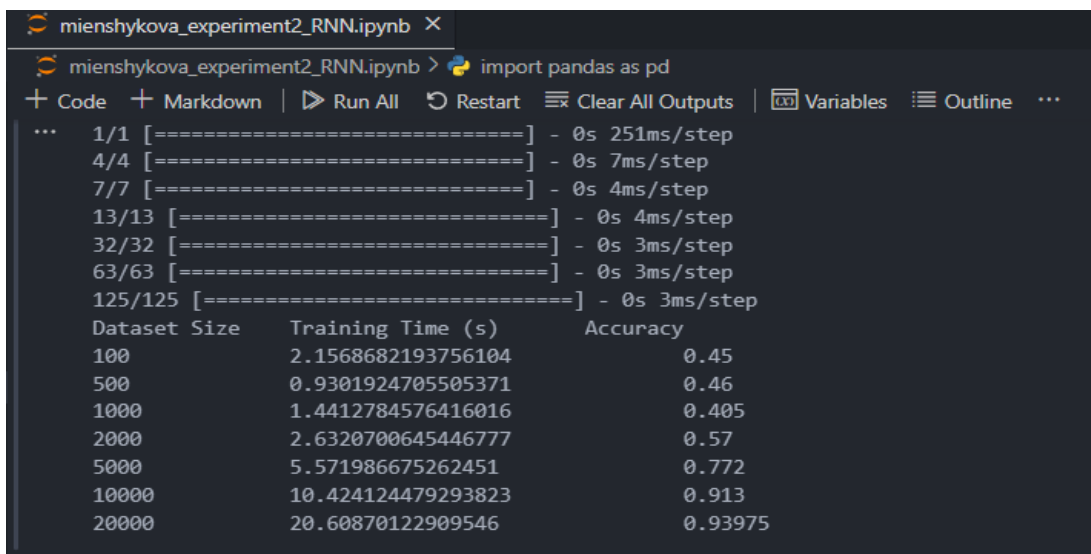


Рис. 11. Вплив розміру набору даних на час навчання та точність моделі RNN

```

mienshykova_experiment2_LSTM.ipynb > import pandas as pd
Code | Markdown | Run All | Clear All Outputs | Outline
... 1/1 [=====] - 1s 538ms/step
4/4 [=====] - 0s 7ms/step
7/7 [=====] - 0s 9ms/step
13/13 [=====] - 0s 9ms/step
32/32 [=====] - 0s 9ms/step
63/63 [=====] - 1s 8ms/step
125/125 [=====] - 1s 9ms/step
Dataset Size  Training Time (s)  Accuracy
100           3.6773722171783447  0.2
500           1.4682366847991943  0.33
1000          3.0088021755218506  0.285
2000          4.755644798278809   0.375
5000          11.317625522613525  0.52
10000         22.192787170410156  0.9295
20000        44.18541717529297  0.955

```

Рис. 12. Вплив розміру набору даних на час навчання та точність LSTM

```

mienshykova_experiment2_GRU.ipynb > import pandas as pd
Code | Markdown | Run All | Clear All Outputs | Outline
... 1/1 [=====] - 1s 501ms/step
4/4 [=====] - 0s 5ms/step
7/7 [=====] - 0s 5ms/step
13/13 [=====] - 0s 5ms/step
32/32 [=====] - 0s 5ms/step
63/63 [=====] - 0s 5ms/step
125/125 [=====] - 1s 5ms/step
Dataset Size  Training Time (s)  Accuracy
100           3.5409343242645264  0.25
500           1.2912118434906006  0.33
1000          1.790619134902954   0.285
2000          3.391169548034668   0.36
5000          8.133849143981934   0.332
10000         16.00895071029663   0.3215
20000        32.01031470298767   0.7185

```

Рис. 13. Вплив розміру набору даних на час навчання та точність моделі GRU

Точність моделі на тестовому наборі даних варіювалася залежно від розміру навчального набору. Модель досягала відносно низької точності для невеликих наборів даних, наприклад, при використанні 100 зразків. Однак із збільшенням розміру набору даних точність покращувалася, досягаючи піку приблизно при 20 000 зразках. Після досягнення певного розміру набору даних покращення точності може зупинитися або навіть почати зменшуватися, що

свідчить про потенційне зменшення прибутковості від подальшого збільшення обсягу даних.

Загалом, модель GRU продемонструвала типову поведінку у відповідь на зміну розміру набору даних, із збільшенням часу навчання та варіаціями у точності. Результати підкреслюють необхідність збалансованого підходу до вибору розміру набору даних, враховуючи як можливі переваги від підвищення точності, так і пов'язані з цим обчислювальні витрати.

Аналіз отриманих результатів. Результати експериментів (табл. 2) показують, що з трьох досліджуваних моделей GRU демонструє найвищу точність, за нею йде LSTM, тоді як RNN має найнижчу точність. Модель GRU також має найвищі значення Precision, Recall та F1-Score, що свідчить про її високу якість розпізнавання емоцій. LSTM показує хороші результати, проте поступається GRU за цими показниками. Модель RNN, навпаки, демонструє найнижчі результати серед трьох моделей, що вказує на її менш ефективне розпізнавання емоцій порівняно з LSTM та GRU.

Модель RNN має найнижчу точність та ефективність серед усіх моделей, але відзначається найменшим часом навчання. Це може бути корисним, якщо швидкість навчання є пріоритетом, хоча за це доводиться платити точністю результатів. Модель LSTM добре справляється із завданням емоційного аналізу, досягаючи значно кращих результатів у порівнянні з RNN,

але трохи поступається GRU. Модель GRU забезпечує найкращі результати з точки зору точності, precision, recall та F1-Score, що робить її найпридатнішою для завдань емоційного аналізу, хоча вона може вимагати більших обчислювальних ресурсів.

Отже, GRU є найкращим варіантом для емоційного аналізу завдяки високим значенням точності та інших метрик. LSTM також може бути ефективним вибором, особливо якщо ресурси обмежені, тоді як RNN, незважаючи на швидке навчання, не є оптимальним вибором через низьку точність.

Тест масштабованості (табл. 3) показав, що всі три моделі демонструють збільшення часу навчання зі збільшенням розміру набору даних. Модель RNN має найменший час навчання для найменших наборів даних, але виявляється неефективною для більших обсягів. LSTM показує найкращі результати точності для всіх розмірів набору даних, але потребує більше часу

Таблиця 2

Експеримент 1 – Аналіз емоцій

Модель	Точність (Аccuracy)	Точність (Precision)	Пригадування (Recall)	F1-score
RNN	0.22099	0.04883	0.22099	0.07999
LSTM	0.73768	0.75949	0.73768	0.73326
GRU	0.82222	0.82882	0.82222	0.82245

Таблиця 3

Експеримент 2 – Тест масштабованості

Модель	Розмір набору даних	Час навчання (с)	Точність
RNN	100	3.68	0.2
	500	1.47	0.33
	1000	3.01	0.285
	2000	4.76	0.375
	5000	11.32	0.52
	10000	22.19	0.9295
	20000	44.19	0.955
LSTM	100	3.68	0.2
	500	1.47	0.33
	1000	3.01	0.285
	2000	4.76	0.375
	5000	11.32	0.52
	10000	22.19	0.9295
	20000	44.19	0.955
GRU	100	3.54	0.25
	500	1.29	0.33
	1000	1.79	0.285
	2000	3.39	0.36
	5000	8.13	0.332
	10000	16.01	0.3215
	20000	32.01	0.7185

для навчання. Модель GRU також демонструє хороші результати, особливо при середніх та великих розмірах даних, хоча її точність може стабілізуватися або навіть зменшуватися при дуже великих наборах даних.

Модель RNN показує швидке навчання для невеликих обсягів даних, але її точність значно поступається LSTM та GRU, що робить її менш придатною для завдань з великими обсягами даних. LSTM залишається найкращим вибором для роботи з великими наборами даних, оскільки вона показує високу точність, хоча час навчання також збільшується. GRU є прийнятним варіантом для середніх та великих наборів даних, демонструючи більш ефективний час навчання порівняно з LSTM при прийнятних результатах точності.

Таким чином, вибір між LSTM та GRU може залежати від конкретного завдання, обсягів даних та доступних ресурсів. LSTM виявляється більш універсальною моделлю з високою точністю для різних обсягів даних, але важливо враховувати час навчання та ресурси при її виборі.

На обмежених ресурсах моделі демонструють різні результати. RNN досягає досить високої точності на низьких ресурсах і має найменший час навчання, що може бути перевагою на обмежених пристроях. LSTM показує високу точність на обмежених ресурсах, хоча її час навчання є вищим порівняно з RNN, але вона залишається працездатною на низьких ресурсах. GRU, хоча і швидша за LSTM, демонструє найнижчу точність серед трьох моделей на обмежених ресурсах, що вказує на втрату точності при швидкому навчанні.

Отже, LSTM є найефективнішою моделлю для використання на обмежених пристроях, зберігаючи високу точність. RNN може бути вибором у випадках, коли основний критерій – це швидкий час навчання. GRU, хоч і швидший за LSTM, але втрачає в точності на низьких ресурсах, що робить LSTM більш придатною для ефективного використання на обмежених пристроях.

Висновки. У ході експериментального дослідження було проведено аналіз моделей глибокого навчання – RNN, LSTM та GRU – в контексті обробки природної мови (NLP) за різними критеріями: емоційний аналіз, масштабованість та ефективність на обмежених ресурсах.

Перший експеримент, присвячений емоційному аналізу, показав, що модель GRU досягла найвищих показників *accuracy*, *precision*, *recall* та *F1-Score*. LSTM також продемонструвала ефективність, але трохи поступилася GRU за

цими метриками. Модель RNN виявила найнижчу точність та загальну ефективність серед трьох досліджуваних моделей.

У другому експерименті, що тестував масштабованість моделей, було виявлено, що RNN є швидкою для роботи з невеликими наборами даних, але неефективною для більших обсягів. LSTM продемонструвала найвищу точність на великих наборах даних, залишаючись оптимальною моделлю для їх обробки. GRU показала хороші результати, особливо з середніми та великими розмірами даних.

У третьому експерименті, який перевіряв ефективність моделей на обмежених ресурсах, LSTM зберегла високу точність, підтвердивши свою ефективність у таких умовах. RNN показала відносно непогану точність і мала найменший час навчання, що може бути перевагою на пристроях з обмеженими ресурсами. GRU, хоч і працювала швидше, але втратила точність у порівнянні з LSTM та RNN на обмежених ресурсах.

Аналіз результатів підтверджує відповідність моделей їхнім перевагам і недолікам. RNN виявилася ефективною в обробці різної довжини послідовностей та моделюванні часових залежностей, проте мала проблеми із зникаючим градієнтом та інтенсивними обчисленнями. LSTM підтвердила свої переваги у моделюванні довгострокових залежностей та стійкості до зашумлених даних, але потребує великих обчислювальних ресурсів. GRU, у свою чергу, показала високу ефективність в обчисленнях і здатність обробляти довгострокові залежності, хоча й може бути менш ефективною в деяких завданнях порівняно з LSTM та має схильність до перенавчання.

Вибір між моделями залежить від конкретного завдання, доступних ресурсів та обсягу даних. LSTM є потужною для моделювання довгострокових залежностей, RNN може бути швидким варіантом для невеликих наборів даних, а GRU є ефективною для обмежених ресурсів.

Для емоційного аналізу рекомендовано використовувати GRU, оскільки вона показала найвищу точність і інші метрики. Якщо ресурси обмежені, LSTM може бути хорошим варіантом, оскільки забезпечує високу точність при менших витратах на обчислення. У завданнях, пов'язаних з великими обсягами даних, LSTM є оптимальним вибором завдяки своїй здатності до масштабування та високій точності. GRU також є хорошим варіантом для роботи із середніми та великими наборами даних, але слід враховувати можливі втрати точності при дуже великих наборах. Якщо ресурсів небагато,

LSTM забезпечує високу точність на обмежених пристроях, тоді як GRU може бути використана для більш швидкого навчання на середніх наборах даних.

Для подальших досліджень рекомендується проведення експериментів з іншими моделями глибокого навчання та тестування цих моделей на різних завданнях NLP, таких як машинний переклад або тематичне моделювання. Інтеграція з іншими технологіями, як-от автоматичне збирання даних або високопродуктивні обчислювальні системи, може підвищити ефективність моделей. Дослідження впливу складних даних на продуктивність, а також оптимізація моделей для роботи на обмежених пристроях можуть стати важливими напрямками для подальших експериментів.

Таким чином, LSTM є найбільш універсальною моделлю для роботи з великими обсягами даних та обмеженими ресурсами, забезпечуючи високу точність. GRU може бути ефективною для завдань з обмеженими ресурсами, проте слід враховувати можливу втрату точності. RNN може бути швидким рішенням для невеликих обсягів даних, але не є оптимальною моделлю для великих задач NLP.

Ці дослідження є важливою точкою для подальшого вивчення та оптимізації моделей глибокого навчання для NLP-завдань. Наступні кроки включають розширення обсягу даних, оптимізацію гіперпараметрів та випробування моделей на інших завданнях обробки природної мови для досягнення більшої ефективності в реальних умовах.

ЛІТЕРАТУРА:

1. Natural Language Processing – DeepLearning.AI. URL: <https://www.deeplearning.ai/resources/natural-language-processing/> (дата звернення: 05.09.2024).
2. Afanasieva I., Golian N., Hnatenko O., Daniil Y., Onyshchenko K. Data exchange model in the Internet of Things concept. *Telecommunications and Radio Engineering*. New York, 2019. Vol. 78, № 10. P. 869–878.
3. Golian, V., Golian, N., Afanasieva, I., Halchenko, K., Onyshchenko, K., Dudar, Z. Study of Methods for Determining Types and Measuring of Agricultural Crops due to Satellite Images. 32nd International Scientific Symposium Metrology and Metrology Assurance, MMA 2022, Sozopol, Bulgaria.
4. Recurrent Neural Networks (RNN) – EDUCBA. URL: <https://www.educba.com/recurrent-neural-networks-rnn/> (дата звернення: 05.09.2024).
5. Anish Nama. Understanding LSTM Architecture, Pros and Cons, and Implementation. Medium, 2020. URL: <https://medium.com/@anishnama20/understanding-lstm-architecture-pros-and-cons-and-implementation-3e0cca194094> (дата звернення: 03.09.2024).
6. Anish Nama. Understanding Gated Recurrent Unit (GRU) in Deep Learning. Medium, 2020. URL: <https://medium.com/@anishnama20/understanding-gated-recurrent-unit-gru-in-deep-learning-2e54923f3e2> (дата звернення: 02.09.2024).
7. Pandey P. Emotion dataset. Kaggle. URL: <https://www.kaggle.com/datasets/parulpandey/emotion-dataset> (дата звернення: 05.09.2024).
8. Afanasieva, I., Golian, N., Golian, V., Khovrat, A., Onyshchenko, K. Application of Neural Networks to Identify of Fake News. *Computational Linguistics and Intelligent Systems (COLINS 2023): 7th International Conference, Kharkiv, 20 April – 21 April 2023: CEUR workshop proceedings, No. 3396*. P. 346–358. URL: <https://ceur-ws.org/Vol-3396/paper28.pdf> (дата звернення: 02.09.2024).
9. Turuta O., Afanasieva I., Golian N., Golian V., Onyshchenko K., Suvorov D. Audio processing methods for speech emotion recognition using machine learning. *MoMLet-2024: 6th International Workshop on Modern Machine Learning Technologies*, 31 травня – 1 червня 2024 р., Львів-Шацьк, Україна. С. 75–108.
10. James, G., Witten, D., Hastie, T., Tibshirani, R. *An Introduction to Statistical Learning with Applications in R*. New York: Springer, 2013. 426 p.

REFERENCES:

1. DeepLearning.AI. (n.d.). Natural Language Processing. Retrieved from: <https://www.deeplearning.ai/resources/natural-language-processing/>
2. Afanasieva, I., Golian, N., Hnatenko, O., Daniil, Y., & Onyshchenko, K. (2019). Data exchange model in the Internet of Things concept. *Telecommunications and Radio Engineering*, 78(10), 869–878.
3. Golian, V., Golian, N., Afanasieva, I., Halchenko, K., Onyshchenko, K., & Dudar, Z. (2022). Study of methods for determining types and measuring of agricultural crops due to satellite images. In *Proceedings of the 32nd International Scientific Symposium Metrology and Metrology Assurance (MMA 2022)*, Sozopol, Bulgaria.

4. EDUCBA. (n.d.). Recurrent Neural Networks (RNN). Retrieved from: <https://www.educba.com/recurrent-neural-networks-rnn/>
5. Nama, A. (2020). Understanding LSTM Architecture, Pros and Cons, and Implementation. Medium. Retrieved from: <https://medium.com/@anishnama20/understanding-lstm-architecture-pros-and-cons-and-implementation-3e0cca194094>
6. Nama, A. (2020). Understanding Gated Recurrent Unit (GRU) in Deep Learning. Medium. Retrieved from: <https://medium.com/@anishnama20/understanding-gated-recurrent-unit-gru-in-deep-learning-2e54923f3e2>
7. Pandey, P. (n.d.). Emotion dataset. Kaggle. Retrieved from: <https://www.kaggle.com/datasets/parulpandey/emotion-dataset>
8. Afanasieva, I., Golian, N., Golian, V., Khovrat, A., & Onyshchenko, K. (2023). Application of neural networks to identify fake news. In Computational Linguistics and Intelligent Systems (COLINS 2023): 7th International Conference, Kharkiv, 20 April – 21 April 2023, CEUR Workshop Proceedings (Vol. 3396, pp. 346–358). Retrieved from: <https://ceur-ws.org/Vol-3396/paper28.pdf>
9. Turuta, O., Afanasieva, I., Golian, N., Golian, V., Onyshchenko, K., & Suvorov, D. (2024). Audio processing methods for speech emotion recognition using machine learning. In Proceedings of the 6th International Workshop on Modern Machine Learning Technologies (MoMLet-2024), Nay 31 – June 1, 2024, Lviv-Shatsk, Ukraine (pp. 75–108).
10. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. New York: Springer.