ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

УДК 004.8 DOI: https://doi.org/10.30837/2522-9818.2024.2.166

А. Ховрат

ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ ВИКОРИСТАННЯ ГІБРИДНИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ ВИЯВЛЕННЯ СФАЛЬСИФІКОВАНОЇ АУДІОІНФОРМАЦІЇ В СОЦІАЛЬНО ОРІЄНТОВАНИХ СИСТЕМАХ

Предметом дослідження є проблема виявлення фальсифікованої інформації, зокрема в аудіоформаті, у соціально орієнтованих системах. **Мета роботи** – розроблення ефективної моделі для визначення факту підроблення звукових даних, яка основана на рекурентних і згорткових нейронних мережах із використанням технології *MapReduce* для паралелізації. У статті розв'язуються такі **завдання**: визначення особливостей аудіо в соціально орієнтованих системах; аналіз алгоритмів для передоброблення аудіоінформації як у перетвореному на текст вигляді, так і у вигляді сигналу; формування переліку цільових архітектур нейромереж та розкриття особливостей їх імплементації; експериментальна перевірка ефективності обраних підходів. Упроваджуються такі **методи**: аналітичний та індуктивний – для визначення цільового набору архітектур нейронних мереж; експертне оцінювання – для формування найбільш впливових факторів ефективності; експериментальний, багатокритеріального оцінювання та статистичні методи аугментації інформації – для визначення найбільш ефективної моделі. **Досягнуті результати**. Сформовано алгоритм передоброблення аудіоінформації для можливості застосування рекурентних і згорткових мереж. Імплементовано декілька підходів до класифікації інформації з використанням аугментації, основаної на векторній авторегресії та технології паралелізації *MapReduce*. Визначено, що найбільш ефективною моделлю, за сформованою задачею багатокритеріального вибору, є поєднання двоспрямованої рекуретної нейромережі з підтримкою короткочасної та довгострокової пам'яті із декількома згортковими мережами. Показано переваги використання технології *MapReduce* для оптимізації часу навчання й передоброблення інформації та визначено набір відкритих питань для подальшого дослідження й прикладного впровадження. **Висновки**. Застосування аналітичного та індуктивного підходу з подальшим проведенням експериментальної перевірки дало змогу розробити ефективний (з точністю понад 96%) механізм виявлення сфабрикованої інформації як у вигляді сигналу, так і у текстовій формі. Досягнутий результат дає підстави стверджувати про доцільність упровадження запропонованого підходу, що зменшує вплив подібної інформації в соціально орієнтованих системах, особливо під час кризових явищ.

Ключові слова: аугментація сигналів; векторна авторегресія; класифікація; оброблення природних мов; фейкова інформація.

Вступ

Упродовж останніх десятиліть технології, спроможні сфальсифікувати інформацію, досягли того рівня, коли про необхідність виявлення підробок у соціально орієнтованих системах говорять на законодавчому рівні [1]. Варто зазначити, що ступінь гостроти проблеми диверсифікується залежно від виду відповідних даних. Зокрема щодо відеоінформації спотворення ще не змогло досягти необхідного рівня [2]. Якщо йдеться про текст і фото, то вже існують ґрунтовні дослідження та навіть певні їх імплементації для визначення факту підробки [3, 4]. Водночас фальсифікація аудіо лише нещодавно змогла перейти межу простої ідентифікації, тобто з використанням людського слуху. Подібний стан речей дає змогу пересічним громадянам сплутати фейковий запис із реальним.

У звичайних умовах проблема може породити локальні конфлікти в групі людей, особливо це помітно в соціальних мережах [5]. Гострішою ситуація стає в умовах військово-політичної нестабільності, коли будь-яка інформація сприймається крізь призму інтенсифікованих емоцій, що сповільнюють процес критичного мислення. У разі, коли підробки є частиною новинного інформаційного поля, вони можуть прискорити соціальні зрушення, викликані кризою, і таким чином посилити її наслідки [6]. Це може мати економічний характер, суто соціальний або навіть військовий, і зрештою негативно позначитися на настроях населення. Як приклад подібної ситуації можна згадати фейкові аудіо, пов'язані з пандемією COVID-19 [7], чи величезну кількість сфальсифікованих записів на початку вторгнення Російської Федерації на територію України [8], що використовувалися

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

для приховання фактів порушення законів ведення війни чи дискредитації Збройних сил України. Необхідно зауважити, що, хоча можливості для якісного підроблення аудіоінформації з'явилися відносно нещодавно, їх вже значно простіше реалізовувати, на відміну від фото, яке потребує достатньо часу й значного обсягу вихідного матеріалу. Це зі свого боку лише посилює можливі ризики для суспільства й держави.

Аналіз останніх досліджень і публікацій

У дослідженні фейкових текстових новин декілька груп іспанських учених [9, 10] показали, що машинне навчання за своєю сутністю потребує достатньої кількості інформації для досягнення позитивного (точність понад 95%) результату класифікації. До того ж таке навчання є доволі чутливим до інформаційних викидів. Однак, зважаючи на наявні у відкритому доступі бази даних, зазначені недоліки не є значними, що було доведено в праці українських дослідників [11]. Подібне виправдання можна застосувати й для інформації інших видів, зокрема аудіо. Це продемонстрували науковці з Массачусетського технологічного інституту в праці [12].

Щодо інших способів виявлення сфальсифікованої інформації, то ще одним доволі популярним методом є створення графових моделей, яке було широко досліджене вченими з Гарварду [13] на прикладі підроблених акаунтів. Зазначений метод гарантує швидкий результат за мінімальних базових даних. Однак у застосуванні для аудіо- чи текстової інформації метод вимагатиме значного передоброблення, і таким чином виграш у швидкодії нівелюється. Якщо ж розглядати питання виявлення неправдивої інформації, то варто згадати й про проблему виявлення спаму. Групою китайсько-американських учених доведена можливість ефективного застосування марковських мереж [14]. Однак, беручи до уваги особливість галузі, їх застосування є доволі громіздким і вимагатиме значних обчислювальних потужностей, що довів канадський дослідник з Монреалю [15].

Під час розгляду візуальної інформації, зокрема фото чи відео, можна застосовувати авторегресійні моделі для виявлення відхилень від базисних значень, інакше кажучи, для виявлення факту маніпуляції з даними. Цей підхід був успішно застосований ученими зі Стенфорду. Водночас його також можна використати для дослідження аудіоінформації у вигляді сигналу. Однак варто зауважити, що спосіб є найбільш ефективним, коли в оригінальний запис було поміщено фрагмент підробки або ж замінена послідовність сигналу.

Іншою можливістю є застосування авторегресії для виявлення факту синтезу інформації (лише за наявності оригінальних записів цільової особи). В умовах контекстуального викривлення подібні моделі не дадуть бажаного результату. Саме тому надалі вони не розглядатимуться як класифікаційні. Натомість, як показала група китайських дослідників, авторегресію можна застосувати для аугментації [16]. Це дасть змогу уникнути проблеми нестачі інформації. Хоча тут варто зауважити, що в процесі розгляду довгострокового проміжку, особливо пов'язаного із соціальними катастрофами та іншими виплесками соціальної активності, зазначений підхід потребуватиме суттєвого обсягу інформації про зовнішні показники. Подібну проблему описують представники Корнелівського університету [17], наголошуючи, що під час розгляду завдання з обмеженою зовнішньою інформацією точність цього підходу (як для прогнозування, так і генерації) суттєво падає. Тому в межах цієї роботи розглядатимуться лише нетривалі (до двох місяців) хронологічні межі.

Визначення не розв'язаних раніше частин загальної проблеми. Мета роботи, завдання

Зазначений стан речей став підґрунтям для того, що у все більшої кількості країн обговорюють боротьбу зі сфабрикованою інформацією, хоча переважно й звертають увагу лише на її текстовий складник. Натомість питання виявлення видозміни аудіо є доволі відкритим, хоча й частково розглянутим, особливо коли йдеться про доповнення реальної інформації, а не синтетичну генерацію. З огляду на світову практику найбільш ефективними в цьому разі є нейронні мережі. Зважаючи на міжнародний досвід, вирішено зосередитися на згорткових та рекурентних нейронних мережах, однак розглядати не їх базові варіації, а більш просунуті та адаптовані для роботи зі значними обсягами текстової інформації – двоспрямовані рекурентні мережі з підтримкою довгострокової та короткочасної пам'яті (*BiLSTM*) і гібридні згортково-рекурентні мережі.

Однак необхідно зауважити, що вказані моделі потребують значних обсягів вихідної інформації,

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

а також часу для навчання моделі. Для розв'язання зазначених проблем можна використовувати принципи аугментації аудіоматеріалів та паралелізму. Однак класичні форми розпаралелювання не гарантують значного виграшу у швидкості навчання моделі [18]. Як базову альтернативу зазначеним принципам використовують технологію *MapReduce*, що, окрім навчання, дає змогу пришвидшити й безпосередньо визначити факт фальсифікації. У межах цієї роботи розглянуто моделі класифікації фейкових аудіо на основі нейронних мереж і можливості модифікації цих алгоритмів і засобів їх розпаралелювання.

Мета роботи – розроблення ефективної моделі для визначення факту підроблення звукової інформації з використанням технології *MapReduce*. Для досягнення окресленої мети сформовано такий перелік *завдань*:

– визначити особливості аудіо в соціально орієнтованих системах (надалі для спрощення позначатимемо їх як новини);

– описати алгоритми, що дало б змогу передобробити аудіоінформацію як у перетвореному на текст вигляді, так і у вигляді сигналу;

– дослідити обрані архітектури нейронних мереж і визначити їх основні гіперпараметри;

– розкрити сутність імплементації технології *MapReduce*;

– сформувати план експерименту;

– проаналізувати результати дослідження та сформулювати відповідні висновки на основі розв'язання задачі багатокритеріального вибору.

Матеріали та методи

Сутність викривлення аудіоінформації

Почнемо розгляд з виявлення особливостей фейкової інформації. Насамперед необхідно зауважити, що фальсифікувати аудіо можна по-різному, зокрема такими способами:

– синтетичне створення аудіо за допомогою засобів штучного інтелекту: має місце за наявності значного обсягу вихідного матеріалу та необхідності отримати голос конкретної особи;

– компонування наявних звукових доріжок для викривлення сутності оригінальної інформації: на відміну від попереднього способу, тут можуть не застосовуватися генераційні алгоритми, однак необхідність отримання голосу конкретної особи зберігається;

– контекстуальне викривлення: у разі, коли власник голосу на аудіо не є важливим, на доріжках можуть записуватися неправдиві новини чи оголошення.

Незважаючи на природу фальсифікації, усі випадки зосереджуються на видозміні контексту з метою досягнення необхідного результату: погіршення настрою населення, шантаж тощо. З огляду на це можна сформувати такий перелік способів, як за наявною в повідомленні інформацією виявити та класифікувати неправдиві аудіо:

– використання неприродної кількості риторичних запитань, якщо йдеться про контекстуальне викривлення суспільно важливої інформації. У лінгвістичних дослідженнях зазначено, що в офіційно-діловому та публіцистичному стилях, призначених для ЗМІ, подібний тип мовленнєвих конструкцій майже відсутній [19]. Ця особливість властива як для текстових новин, так і аудіо, відео;

– відсутність заперечувальних конструкцій для зменшення когнітивного навантаження в поєднанні з песимістичним забарвленням обраних слів. Як приклад можна навести заміну слова "проблема" на "катастрофа". Однак необхідно зауважити, що в усному мовленні використання ненормативної лексики не дає змоги напевно визначити характер забарвлення, тому для подальшого аналізу оцінка подібних слів формуватиметься на основі контексту;

– вживання закликів і спонукань у недоречних формах. У разі контекстуального викривлення, мета якого – замінити реальні новини, подібні конструкції одразу вказують на неправдивість та некоректність інформації. Однак унаслідок розгляду мімікрійних записів ці особливості можуть визначити мету зловмисників;

– використання невиправданої кількості займенників. Цей фактор здебільшого необхідний для контекстуального викривлення, що наслідує публіцистичний стиль мовлення.

Зазначені характеристики не є вичерпними, однак маємо наголосити, що в разі оброблення трансформованих у текстовий вигляд аудіозаписів деякі особливості можуть не виявитися. Зокрема відомо, що для формування фейкових новин часто вживають короткі речення та слова, або ж у них наявна значна кількість різноманітних помилок [20]. Ці та подібні ним характеристики не беруться до уваги надалі, оскільки вони можуть виявитися через некоректність розпізнавання аудіо чи загалом бути особливою частиною процесу мовлення

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

людей, присутніх на аудіозаписі (наприклад, за умови змішування двох мов чи використання регіональних діалектизмів). Ці самі особливості можуть зумовити більшу кількість *false positive* випадків у виявленні фейків.

Аналіз аудіо як тексту

Унаслідок аналізу праці групи китайських учених визначено, що створення власного модуля *Speech to Text* супроводжується такими проблемами [16]:

– якість записів для тренування;

– нестача інформації для формування моделі (особливо гостро проблема постає для мов із невеликими корпусами);

– ігнорування дефектів вимови;

– коректність оброблення діалектизмів, неологізмів, скорочень тощо.

Зазначений перелік не є повним, тож аби уникнути вказаних проблем і досягти найліпшого результату перетворення аудіо на текст, вирішено використати *Google Speech to Text*, зокрема його відповідну обгортку для мови програмування *Python 3*. Додатково за допомогою голосу, записаного 20 людьми з різних регіонів України та різними вадами мовлення, а також 20 записами з українськомовних фільмів було встановлено:

– система має обмежені можливості в розпізнаванні скорочень;

– якщо паузи між словами є дуже тривалими, то модуль визначатиме окремі речення;

– якість записів не суттєво впливає на ефективність розпізнавання: наявність додаткового шуму нівелюється завдяки стадії передоброблення аудіо;

– без уваги до вищезазначеного точність розпізнавання сягала понад 95%, винятком стали сільські говори західних і східних областей.

Надалі зазначені твердження вважатимемо обмеженнями цієї статті.

Щоб перетворити добуту текстову інформацію в числове подання, скористаємося таким алгоритмом:

– розбиваємо текст на речення та окремі слова;

– вилучаємо слова без суттєвого лексико-графічного навантаження та ті, що не впливають на результат роботи алгоритмів (так звані стоп-слова). Наприклад: "однак", "це", "або", "тощо";

– формуємо на основі добутих лексем словник тексту;

– виокремлюємо основи кожного слова в словнику та прибираємо повтори (здійснюємо операцію стемінгу);

– визначаємо лему для кожної лексеми в словнику та знову прибираємо повтори (здійснюємо лематизацію);

– визначаємо частотну характеристку кожного слова (її опис буде здійснено нижче) та його емоційного забарвлення за допомогою засобів *Sentiment Analysis,* вбудованих у моделі *nltk* мови програмування *Python 3*;

– модифікуємо оцінку емоційного забарвлення у межах кожного окремого речення на основі правил, установлених раніше;

– агрегуємо й нормуємо частотно-емоційний показник для кожного речення, він слугуватиме цільовим індикатором для подальшого використання нейронних мереж;

– знаходимо показник підозрілості в нормованому вигляді на основі переліку слів, що часто вживаються у фейковій інформації.

Окрім указаних змінних, вхідними величинами також застосовуватимемо такі показники:

– частотно-емоційна характеристика 50 найбільш популярних новин за дату створення аудіозапису. Це дасть змогу взяти до уваги новинне зовнішнє середовище й, відповідно, скоригувати оцінку класифікатора;

– вага повідомлення. Вирішено створити набір даних, у якому аудіо поділятимуться на чотири групи: фрагменти домашнього діалогу, загальні новини, інформація з місця надзвичайних подій, новини особливої важливості. Маркування здійснюватиметься від 1 до 4 відповідно. Значення показника ваги також буде нормованим;

– ступінь надійності конвертації аудіо в текст. Визначатиметься в процесі порівняння реального тексту повідомлення з тим, що був оброблений *Google Speech to Text*, як відношення правильно розпізнаних слів.

Щодо частотної характеристики, то вирішено використати *BM25*, яка є певною модифікацією *TF-IDF*. Для кращого зрозуміння сутності модифікації детальніше розглянемо базову характеристику. Мета *TF-IDF* – зважати на важливість кожного слова в запиті та тексті з огляду на частоту вживання терміна як у певному документі, так і в корпусі загалом. Умовно слово "і" може бути одним із найбільш поширених у конкретному реченні,

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

але воно часто вживається в корпусі загалом, тож матиме меншу значущість у пошуку за цим корпусом. Метод оснований лише на статистиці та рахується доволі швидко, тож і досі залишається популярним для задач, де не потрібні більш складні рішення. У разі *BM25* до *TF* додається насичена частота терміна. Тобто якщо термінологічна одиниця вже має високу частоту, то після певної познаки зростання частоти не матиме значного впливу на оцінку *TF*. *IDF* використовується так само. Додатково наявні два параметри – $k1$ і b, які можна налаштовувати під конкретний корпус. Параметр $k1$ відповідає за насичення частоти, а b – за міру впливу довжини документа на результати.

Вибір *TF-IDF* був зумовлений результатами попередніх досліджень, присвячених аналізу текстових новин [21]. Неточності та обмеження знизили ефективність запропонованих класифікаційних методів. Зокрема однією з проблем виокремлено перенасичення певними термінами, які не можна вважати стоп-словами, наприклад "катастрофа".

З'ясувавши особливості аудіо в текстовому вигляді, перейдемо до розгляду аудіо як сигналу.

Аналіз аудіо як сигналу

Першим етапом у підготовленні аудіо до його оброблення як сигналу є виокремлення вокалізованої частини від тиші. Подібна операція необхідна, адже в першому фрагменті присутні ключові елементи мовлення людей. Одним із загальновживаних способів маркування аудіосигналу є його розбиття на три стани:

– ділянка тиші (S), де відсутня вимова;

– невокалізована ділянка (U), де результуюча форма сигналу має аперіодичний або випадковий характер (має місце в разі, коли голосові зв'язки не вібрують);

– вокалізована ділянка (V), де результуюча форма сигналу є квазіперіодичною (має місце, коли голосові зв'язки суб'єкта мовлення напружені та, відповідно, вібрують).

Щодо поєднання двох перших ділянок, то зауважимо, що існують методи, які б дали змогу розмежувати тишу від невокалізації, однак вони вимагають постійного переналаштування для різного оточення, що в контексті аудіоновин є малоефективною процедурою. Подібна проблема наявна і для методів, які розмежовують вокалізовану ділянку від інших на основі малої енергії. Тому, зважаючи на зазначене твердження і той факт, що шумове оточення новин може різнитися, хоча є відносно стабільним, було вирішено застосувати методи, основані на розподілі. У цьому разі вважатимемо, що сигнал має гауссівську природу. Отже, щоб виокремити необхідну частину аудіо, можна використати функцію відстані Махаланобіса, яка є класифікатором лінійних шаблонів (*LPC*).

Щоб визначити параметри гауссівського розподілу, необхідно детермінувати базисне вікно. Для цього необхідне впровадження методу експертного оцінювання. Було опитано 30 фахівців з оброблення звуку з Харкова, Києва, Дніпра та Відня. Установлено, що оптимальний розмір вікна становить 200 мс. Тепер візьмемо до уваги формулу визначення гауссівського розподілу для одновимірного випадку:

$$g(x)=\frac{1}{\sqrt{2\pi\sigma}}e^{\frac{(x-\mu)^2}{2\sigma^2}}, \qquad (1)$$

де μ – середнє значення;

σ – стандартне відхилення розподілу.

З огляду на це можна визначити такий набір правил:

$$\begin{aligned} P\left[|x-\mu|\le\sigma\right]&\approx 0,68,\\ P\left[|x-\mu|\le 2\sigma\right]&\approx 0,95,\\ P\left[|x-\mu|\le 3\sigma\right]&\approx 0,997. \end{aligned} \qquad (2)$$

Відстань Махаланобіса визначимо за допомогою формули

$$r=\frac{|x-\mu|}{\sigma}. \qquad (3)$$

Беручи до уваги (2) та (3), можна встановити, що з імовірністю 99,7% відстань становить менше ніж 3.

Процес передбачає такі кроки:

– алгоритм поступово аналізує аудіо у вікні 200 мс та визначає стандартне відхилення та середнє значення;

– для кожного наступного вікна обчислюється відстань Махаланобіса з використанням добутих раніше значень;

– якщо відстань перевищує 3, то вважаємо семпл вокалізованим, в іншому разі замінюємо його на порожні значення, фактично вилучаючи.

Нейронна мережа прийматиме на вхід перетворене аудіо, однак цього разу вікно для розбиття на семпли визначатиметься за допомогою крос-валідації. Як зазначалося вище, для уникнення проблем нестачі

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

аудіосигналів вирішено здійснити аугментацію за допомогою авторегресії.

$$\Phi_0 y_t = \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \Theta_0 u_t + \Theta_1 u_{t-1} + \cdots + \Theta_q u_{t-q} \ , \quad (4)$$

де y_t – K-вимірний часовий ряд;

Φ_i, Θ_j – матриці розмірності $K \times K$, $i = \overline{1, p}$, $j = \overline{1, q}$;

u_t – K-вимірний вектор білого шуму з нульовим середнім.

Тут варто зауважити, що, оскільки матриці коефіцієнтів не вироджені, їх легко нормалізувати

$$\Phi_0 \Delta y_t = \Pi y_{t-1} + \Psi_1 \Delta y_{t-1} + \cdots + \Psi_{p-1} y_{t-p+1} + \Theta_0 u_t + \Theta_1 u_{t-1} + \cdots + \Theta_q u_{t-q} \ , \quad (5)$$

де $\Pi = -\left(\Phi_0 - \Phi_1 - \cdots - \Phi_p\right)$;

$\Psi_i = -\left(\Phi_{i+1} + \cdots + \Phi_p\right)$, $i = \overline{1, p-1}$.

Добуті матриці коефіцієнтів за умови загальної кількості невідомих, що потрібно брати до уваги під час генерації, можуть змінюватися залежно від обраної моделі. У межах цього дослідження розглядатимуться такі варіації:

– проста авторегресія;
– сезонна авторегресія;
– авторегресія розподіленого лагу;
– авторегресія рухомого середнього;
– авторегресія інтегрованого рухомого середнього.

У цьому разі єдиним фактором ефективності можна вважати точність аугментації даних. Фактично необхідно гарантувати, що розподіл між вокалізованими та невокалізованими 200 мс семплами не зміниться.

Розглянувши основні засоби препроцесингу аудіо у вигляді тексту та сигналу, опишемо архітектури нейронних мереж, які плануємо використати.

Архітектури нейронних мереж

Обрані архітектури *BiLSTM* та *CNN-RNN* за своєю сутністю є нащадками класичних згорткових і рекурентних нейромереж. Тож щоб краще зрозуміти ці моделі, здійснемо поступовий огляд базових алгоритмів.

Почнемо з *RNN*. Вона містить декілька прихованих шарів, що працюють один за одним. Водночас кожен наступний шар на вхід отримує результат роботи попереднього. Зазначену особливість прийнято називати короткочасною пам'яттю за аналогією з людським мозком.

Усередині прихованого шару поступово обробляється вихідна інформація із використанням

Формально її можна подати таким чином:

в межах від 0 до 1. Модель (4) можна використовувати в процесі розгляду короткочасних періодів, до того ж потрібно гарантувати відсутність або ж несуттєвість зовнішнього впливу. Оскільки в межах цієї роботи вирішено розглянути середньострокову перспективу, необхідно знайти дельту між (4) та прогнозом на попередній період, тобто $\Phi_0 y_{t-1}$. Отже, маємо формулу:

градієнтів. Однак у разі, коли дані мають особливий характер та не є обмеженими за своєю сутністю (як згори, так і знизу), можуть виникати проблеми вибухового чи напівзниклого градієнтів. Це ситуації, коли значення градієнта починає прямувати до нескінченності та 0 відповідно. З огляду на міжнародний досвід під час аналізу текстової інформації (і сигналів також) ця проблема може мати місце. Щоб подолати вказаний недолік, вирішено використати нейронні мережі із підтримкою короткочасної та довгострокової пам'яті (*LSTM*).

Сутність математичного апарату в *LSTM* полягає в поступовому використанні декількох сигмоїд і гіперболічних тангенсів, що дають змогу коригувати значення таким чином, щоб уникнути спрямування як до 0, так і до нескінченності. Пропонуємо детальніше розглянути основні етапи роботи прихованих шарів цієї архітектури.

На першому етапі у *Forget Gate* додаються дві вхідні вагові інформації та здійснюється множення на сигма-функцію активації. Область значень цієї функції обмежує результат виконання вказаного етапу в межах від 0 до 1. Після завершення результат множиться на дані з каналу довгострокової пам'яті.

Другим етапом є *Input Gate*, що здійснює аналогічне множення на сигма-функцію. Однак цього разу результат коригується за допомогою застосування гіперболічного тангенса. Подібна операція дає змогу нівелювати проблему спрямування до нескінченності в процесі додавання до значення з каналу довгострокової пам'яті.

Унаслідок роботи двох зазначених етапів формується стан пам'яті, що разом із вхідною та попередньою вихідною інформацією слугує базисом для формування нового значення короткочасної пам'яті. Цей етап має назву *Output Gate* і є

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

завершальним кроком виконання одного прихованого шару. Сутність полягає в знаходженні гіперболічного тангенса від значення довгострокової пам'яті, який після цього множиться на сигма-функцію від попереднього значення в короткочасній пам'яті.

Хоча вказана архітектура дає змогу уникнути проблеми градієнтів, вона все ж має один істотний недолік під час оброблення природної мови – неможливість зважати на майбутній контекст. Для кращого розуміння наведемо приклад.

Нехай маємо початок речення "*Apple is something that…*". Звичайна *LSTM*-архітектура не зможе визначити, що саме мається на увазі під "*Apple*" – фрукт чи компанія, бо немає інформації про кінець наведеного речення. Для цієї архітектури варіанти "*Apple is something that competitors simply cannot reproduce*" та "*Apple is something that I like to eat*" є ідемпотентними. Для уникнення зазначеної проблеми вирішено використати двоспрямовану рекурентну нейромережу з підтримкою довгострокової та короткочасної пам'яті (*BiLSTM*). Фактично сутність цієї мережі полягає в поєднанні двох *LSTM*, спрямованих у різні напрямки. У цьому разі "допоміжна мережа" дає змогу зважати на контекст для початку речень.

Після відпрацювання двох підмереж результат обох рівнів поєднується, спочатку способом простої конкатенації, а після цього за допомогою лінійних трансформацій. Для того щоб визначити відповідні операції, вирішено провести крос-валідацію, під час якої встановлено, що найкращий результат досягається за умови використання усереднених значень. Аналогічний висновок було зроблено в процесі експертного оцінювання серед 10 осіб, що займаються обробленням природної мови.

Розглянувши першу із запропонованих архітектур, перейдемо до *CNN*-архітектури.

Порівняно з попередньою, ця модель не має ні короткочасної, ні довгострокової пам'яті. Натомість вона використовує шар згортки, що дає змогу суттєво зменшити розмірність вихідної інформації. Це особливо ефективно в розпізнаванні образів і визначенні факту фальсифікації зображень чи відео.

Щоб мати змогу побудувати якомога ефективнішу *CNN*-архітектуру, необхідно задати низку гіперпараметрів моделі. Одним із найбільш важливих серед них є розмір фільтра. Це елемент прихованого шару, що здійснює прохід поміж інформацією та виконує згортку. Після проведення крос-валідації встановлено, що для обраного випадку найкращим буде фільтр розмірністю $5\times5\times5$.

Тут варто зауважити, що останнє значення розмірності в нашому дослідженні відповідатиме кількості дескрипторів, що утворюють цільову змінну. Саме тому для аналізу аудіо як тексту розмірність становитиме $5\times5\times5$, однак у разі аналізу аудіо як сигналу розглядатимемо як дескриптори лише стандартне відхилення та середнє значення, тож розмірність фільтра становитиме $5\times5\times2$.

У процесі його проходження поміж даними розташований скалярний добуток між записами фільтра та вхідною інформацією. Це дозволить сформувати активаційну карту, розмірність якої дорівнюватиме кількості використаних фільтрів, інакше – глибині нейромережі. З огляду на кількість факторів, що беруться до уваги для класифікації, було вирішено зупинитися на глибині, рівній 5 та 2 відповідно.

Окрім зазначеного гіперпараметра для *CNN*-архітектури, розглядаються такі характеристики:

– розмір ядра (у процесі крос-валідації було використано ядро в межах від 2 до 5 і встановлено оптимальне значення, що дорівнює 4);

– розмір кроку під час розгляду. Зважаючи на рекомендації, що вказують на небажаність використання кроку понад 3 для тексту, було визначено оптимальне значення, що дорівнює 1;

– беручи до уваги встановлений крок, параметр додавання неістотних нулів не застосовуватиметься;

– з огляду на особливість предметної галузі вирішено не застосовувати параметр зміщення.

Варто зауважити, що кількість шарів згорткової мережі для аналізу текстової інформації має дорівнювати 1.

Проблемою зазначеної архітектури за умови її використання для оброблення природних мов є обмеженість щодо уваги до контексту. Звичайно, проходження фільтра дає змогу зважати на окіл кожного зі слів, однак особливість української мови полягає у великих реченнях. Отже, визначений контекст може розташовуватися поза фільтром *CNN*-моделі. Для уникнення цієї проблеми було вирішено поєднати *RNN* та *CNN*.

Хоча способів подібного поєднання існує декілька, у межах цієї роботи розглядатиметься лише *RCNN*-архітектура, що послідовно використовує дві нейронні мережі. Інакше кажучи, після здійснення згортки результат не лише конкатенується, а надсилається до шару з рекурентною нейронною мережею.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Щоб уникнути окреслених проблем під час розгляду тексту, вирішено застосувати не класичну *RNN*-архітектуру, а згадану вище двоспрямовану рекурентну нейромережу з підтримкою довгострокової та короткочасної пам'яті. Отже, архітектуру *RCNN* можна подати у вигляді, зображеному на рис. 1.

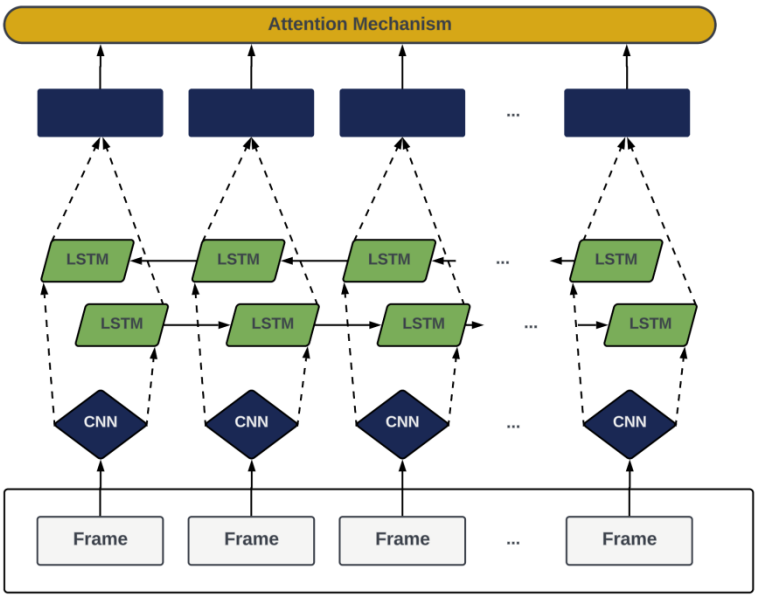

Рис. 1. Схематичне зображення *RCNN*-архітектури

Як зазначалося вище, у використанні складних нейронних мереж є суттєвий недолік – час їх навчання та оброблення. Крім цього, проблемою також є час передоброблення інформації. З метою зменшення впливу вказаних недоліків було вирішено впровадити технологію *MapReduce*.

Сутність технології MapReduce

Технологія *MapReduce* полягає в розподілі вихідного набору інформації таким чином, щоб вона оброблялася на окремих вузлах.

Ключовими операціями є застосування функцій мепінгу та редукції. Перша дає змогу розподілити інформацію між вузлами, на яких здійснюється бажане оброблення, а друга функція натомість збирає дані з усіх вузлів і уніфікує їх.

Варто зауважити, що технологія *MapReduce* визначає лише особливості реалізації відповідних модулів у межах певних фреймворків. Отже, ця реалізація може суттєво відрізнятися. Для цієї роботи обрано технологію *MapReduce* на основі *Hadoop*. Графічно запропоноване рішення можна подати так, як показано нижче (рис. 2).

У цьому разі особливу увагу варто приділити функціям розподілу та комбінування. Вони необхідні для того, щоб здійснити додаткову паралелізацію в кожному з вузлів, застосувавши різні регіони пам'яті. Щоб краще зрозуміти сутність цього підходу, можна вважати базові вузли процесами, а вказані регіони пам'яті – потоками.

Окрім цих двох функцій, важливою особливістю є сортування інформації перед редукцією. У статті розглядається інформація, що суттєво залежить від порядку й не має додаткових часових міток, таких як у часових рядах. Щоб уникнути проблеми внаслідок редукції, вирішено додати поле з порядковим номером кожного фрагменту тексту / сигналу. За ним і здійснюватиметься сортування.

MapReduce використовуватиметься незалежно в процесі препроцесингу вихідної інформації та навчання нейронних мереж.

Для здійснення передоброблення у разі сигналів важливим є лише здійснення редукції в правильному порядку. Для оброблення аудіо як тексту необхідно брати до уваги важливість формування якомога більшого словника. Для цього вирішено створити окрему нереляційну базу даних із підтримкою багатопотоковості, куди після базового оброблення (вилучення стоп-слів, лематизації, стемінгу) записуватиметься весь наявний словник. Отже, що більше матеріалу буде оброблено, то вищою буде точність формування відповідної частотної характеристики.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

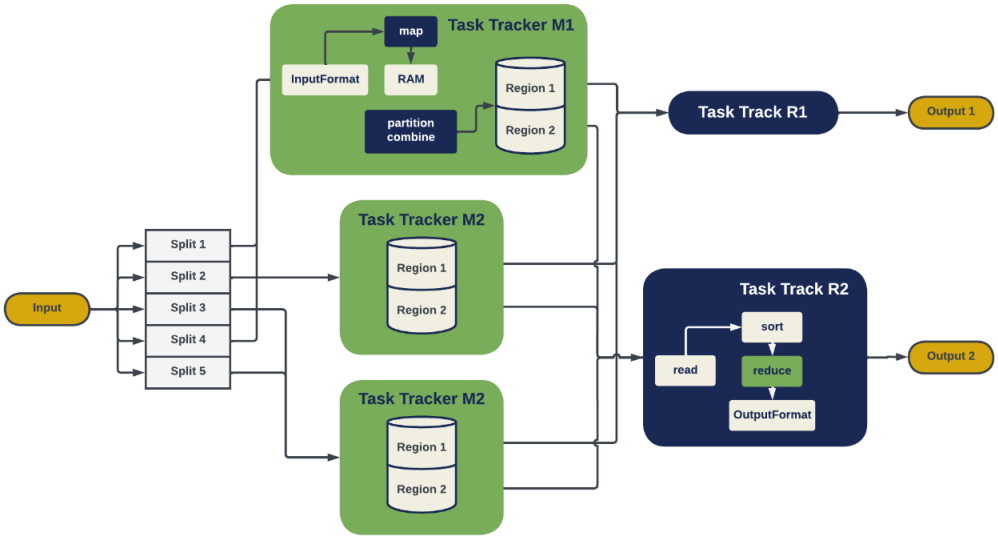

Рис. 2. Схематичне зображення *MapReduce* на основі *Hadoop*

Для *RCNN*-архітектури першим кроком є шар із *CNN*. У ньому ітеративно регулюються вагові коефіцієнти, обчислюючи їх часткові градієнти після того, як кожен набір навчальної інформації поширюється з допомогою мережі.

Так, розпаралелювання під час фази навчання може бути здійснено способом розподілу даних на декілька фрагментів. Потім кожен фрагмент передається в кілька *CNN*, і всі *CNN* навчаються незалежно. Після цього результати агрегуються за допомогою редуктора для отримання остаточної інформації, що потім використовується для оновлення вагових коефіцієнтів для подальшої ітерації.

Після завершення роботи шару *CNN* агрегована інформація передається у *BiLSTM*. Щоб пришвидшити двоспрямовану нейромережу, можна розподілити роботу двох нейронних мереж між двома вузлами. У цьому разі функція редукції фактично слугуватиме функцією агрегації результатів двох мереж.

Серед загальних переваг запропонованого підходу можна виокремити його масштабованість, відносну дешевизну, простоту у використанні та можливість моніторингу виконання за допомогою відповідних засобів *Hadoop* (якщо потрібен внутрішній моніторинг, використовуються базові методи мови програмування *Python*). Серед недоліків можна наголосити на необхідності створення великого обсягу програмного коду, прихованості оброблення (хоча і з можливістю перегляду лог-файлів) та потребі в тривалому налаштуванні конфігурації.

Експериментальне середовище

Зважаючи на особливість запропонованого дослідження, обрано метод контрольованого експерименту. Базове середовище виконання має такий набір характеристик:

- CPU: Intel Core i5-1135G7;
- RAM: 16 Гб;
- VRAM: 4 Гб;
- ОС: Ubuntu 21.04.

Перелічені характеристики майже в повному обсязі продубльовані на віртуальних вузлах, на яких планується здійснюватися часткове обчислення (було зменшено *RAM* з 16 до 8). Їх кількість становитиме від 3 до 4 (2 – у разі паралелізації двоспрямованих нейронних мереж).

Засобом обчислення часу виконання обрано бібліотеку *datetime* для *Python* 3, що має точність до наносекунди. Щоб базові обчислення не сповільнювали роботу програми, застосовано бібліотеки *numpy* та *polars*. Для оброблення природних мов (зважаючи на лематизацію, токенізацію та інші необхідні функції) обрано *python*-версію бібліотеки *nltk*. З метою реалізації нейронних мереж використано *tensorflow* із інструментарієм, що надає *pipeline* субмодуль.

У процесі розгляду аудіо у вигляді тексту, як уже зазначалося, було вирішено застосувати *Google Cloud Platform* для уникнення впливу оточення на швидкість роботи. Оптимізація цього інструменту дала змогу зменшити час затримки в разі отримання розшифровки текстової інформації з 10 с (на оброблення двохвилинного аудіо) до 2 с.

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

Перше значення отримане за допомогою власноруч побудованого інструменту розпізнавання мови.

Інформацією для перевірки використано власноруч сформовані набори аудіо, згенеровані на основі текстових новин та частково видозмінених різними способами, описаними раніше.

Перший набір інформації стосується повномасштабного вторгнення Росії на територію України та містить як власне новини, так і реакції на певні події користувачів соціальних мереж, експертів із телебачення тощо. Другий набір інформації присвячений виборчому процесу 2019 р., що супроводжувався появою значної кількості неправдивих відомостей. Кожна з цих вибірок ділитиметься у співвідношенні 80 до 20 на навчальну та тестову підвибірки відповідно.

Щоб порівняти різні види нейронних мереж з та без використання *MapReduce,* необхідно визначитися з основними критеріями вибору. З огляду на те, що розглядається завдання класифікації для соціально гострих процесів, найбільш важливими критеріями є економія часу й точність класифікації.

Загалом було обрано такий перелік факторів:

– показник точності;

– економія часу навчання моделі за однакових потужностей;

– економія часу передоброблення інформації;

– економія мінімально допустимого обсягу інформації для досягнення *Accuracy* = 90%;

– можливість уваги до контексту.

Тут варто зауважити, що показник економії часу передоброблення інформації є важливим лише для визначення авторегресійного алгоритму та виграшу у швидкості, що надає технологія *MapReduce*. Тож для оцінювання ефективності нейронних мереж використовуватиметься лише чотири метрики з наведених вище.

Визначившись із критеріями, опишемо відповідні шкали оцінювання.

Економія часу навчання моделі вимірюватиметься в секундах за допомогою зазначеної вище бібліотеки. Сам показник у цьому разі не обмежуємо. Щоб зменшити вплив можливого вимірювання, викликаного проблемами з точністю роботи часових модулів чи оточення, вирішено проводити по п’ять замірів для показників часу та перевірити точність прогнозування на двох вибірках інформації.

Точність класифікації визначена за допомогою комбінації *F1-score* та *Precision*, нормалізована в межах від 0 до 1. Замір точності здійснюватимемо для двох вибірок і братимемо усереднене значення.

Як це було продемонстровано на прикладі *LSTM*- та *BiLSTM*-архітектур, на контекст можна зважати в різному обсязі:

– повною мірою в обох напрямках – 5 балів;

– лише в одному напрямку – 4 бали;

– лише в околі в обох напрямках – 3 бали;

– лише в околі в одному напрямку – 2 бали;

– не береться до уваги – 1 бал.

Варто зауважити, що в класифікації сигналів цей показник не використовуватиметься.

Щоб визначити, яка з моделей є найбільш ефективною за поданими вище критеріями, застосовано принцип лінійної адитивної згортки з ваговими коефіцієнтами. Для визначення вагових коефіцієнтів проведено експертне оцінювання серед журналістів та аналітиків (кількість рецензентів становила 50 осіб). Отже, перейдемо до визначення вагових коефіцієнтів. У питанні класифікації як сигналів, так і тексту найбільш важливим є показник точності. На другому місці – можливість зважати на контекст, на третьому – показник часу. Отже, можемо призначити:

– для точності – 16 очок;

– для можливості зважати на контекст – 10 очок;

– для економії часу навчання моделі за однакових потужностей – 2 очки;

– для економії мінімально допустимого обсягу інформації з метою досягнення потрібного рівня *Accuracy* – 2 очки.

З огляду на це отримуємо такі вагові коефіцієнти для кожного критерію:

– для точності: $16/30 = 8/15$ у разі аудіо як тексту; $16/20 = 0.8$, якщо аудіо у вигляді сигналу;

– для можливості зважати на контекст: $10/30 = 5/15$, якщо аудіо – текст;

– для економії часу навчання моделі за однакових потужностей: $2/30 = 1/15$ у разі аудіо як тексту, $2/20 = 0.1$, якщо аудіо у вигляді сигналу;

– для економії мінімального допустимого обсягу інформації: $2/30 = 1/15$ у разі аудіо у вигляді тексту, $2/20 = 0.1$, якщо аудіо як сигнал.

Наступним важливим елементом експериментального середовища є визначення можливих похибок. З огляду на описаний план можна виокремити такі фактори, що здатні вплинути на результат:

– під час перевірки економії часу – людський фактор та інструментальна похибка;

– під час перевірки точності – проблема даних.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Щоб пом'якшити зазначені невизначеності показники вимірюватимуться декілька разів.

Результати дослідження

Спочатку подамо значення можливості уваги до контексту для кожної з наведених вище моделей:

– *CNN* – 3 бали, контекст береться до уваги лише в околі в обох напрямках за допомогою функції згортки;

– *RNN* – 2 бали, контекст береться до уваги лише в околі в одному напрямку за допомогою короткочасної пам'яті;

– *LSTM* – 4 бали, контекст береться до уваги лише в одному напрямку;

– *BiLSTM* – 5 балів, контекст береться до уваги повною мірою в обох напрямках унаслідок двоспрямованості мережі;

– *RCNN* – 5 балів, контекст береться до уваги повною мірою в обох напрямках за допомогою функцій згортки та двоспрямованої мережі з довгостроковою пам'яттю.

Почнемо з показника економії часу передоброблення інформації в разі використання сигналів (з аугментацією за допомогою авторегресійних моделей).

Результати наведені в табл. 1. Усі значення економії пораховані відповідно до найповільнішого алгоритму – послідовної версії векторної авторегресії інтегрованого рухомого середнього.

Значення критеріїв для кожного методу налаштування моделі подано в табл. 1. Для спрощення викладок запропоновано такі познаки:

– *R* – класична векторна авторегресія;

– *RS* – сезонна векторна авторегресія;

– *RL* – векторна авторегресія розподіленого лагу;

– *RMA* – векторна авторегресія рухомого середнього;

– *RIMA* – векторна авторегресія інтегрованого рухомого середнього.

Таблиця 1. *Збереження часу передоброблення для сигналу (у мілісекундах)*

Послідовний підхід					MapReduce підхід				
R	**RS**	**RL**	**RMA**	**RIMA**	**R**	**RS**	**RL**	**RMA**	**RIMA**
55	47	36	12	0	169	138	105	49	4
58	45	30	13	0	173	135	101	52	5
61	48	36	20	0	165	130	99	47	3
57	42	35	22	0	168	132	102	49	4
59	48	38	18	0	166	132	101	50	5

Знайдемо середнє значення для кожного випадку за умови послідовних версій. Для *R* маємо 0.058 с; для *RS* – 0.046 с; для *RL* – 0.035 с; для *RMA* – 0.017 с; для *RIMA* – 0 с. Як бачимо, в середньому алгоритми рухомого середнього та інтегрованого рухомого середнього є значно повільнішими. Це пояснюється тим, що вони беруть до уваги екзогенні змінні в повному обсязі і врегульовують шуми. Оскільки для нашого дослідження точність аугментації не є найсуттєвішим показником, з огляду на результати вирішено скористатися класичною векторною авторегресією.

Якщо ж порівнювати з версіями *MapReduce*, то виграш у швидкості становить ~ 2.9 для кожної моделі. Якщо покращити конфігурацію для *MapReduce* і збільшити кількість вузлів до 4, то виграш становитиме ~ 3.74.

Аналіз аудіо як тексту показав різницю лише між паралелізованою версією та послідовною. Для трьох *MapReduce* вузлів прискорення становило ~ 3.1, у разі чотирьох ~ 4.3. Кількість вузлів менша за прискорення через додаткову оптимізацію запитів до *Google Speech to Text API*.

Перейдемо до результатів замірів економії часу навчання та почнемо з аналізу аудіо як сигналу (табл. 2).

Цього разу застосування паралелізації наявне лише для *BiLSTM*- і *RCNN*-архітектур для визначення загального рівня прискорення. Найбільш повільною є модель, побудована на *RCNN*-архітектурі.

Маємо такі середні значення показників: *CNN* – 51 с; *RNN* – 46 с; *LSTM* – 30 с; *BiLSTM* – 16 с; *RCNN* – 0 с. Як бачимо, що складніша за своєю сутністю архітектура, то повільнішим є навчання відповідної моделі.

Прискорення, досягнуте за допомогою *MapReduce* для *BiLSTM,* становило 2 (пояснюється простотою паралелізації та обмеженістю двома вузлами); для RCNN ~ 3.52 (для чотирьох вузлів результат збільшився до ~ 4.68).

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Таблиця 2. *Збереження часу тренування мереж для сигналу (у секундах)*

Послідовний підхід					MapReduce підхід	
CNN	**RNN**	**LSTM**	**BiLSTM**	**RCNN**	**BiLSTM**	**RCNN**
50	48	29	15	0	30	25
49	44	30	14	0	29	24
52	47	28	17	0	31	27
54	45	33	16	0	28	24
50	46	30	18	0	30	26

Унаслідок замірів для навчання обраних нейронних мереж із текстовою інформацією досягнуті такі результати середнього значення економії часу: *CNN* – 45 с; *RNN* – 41 с; *LSTM* – 24 с; *BiLSTM* – 12 с; *RCNN* – 0 с; *BiLSTM based MapReduce* – 24 с; *RCNN based MapReduct* – 22 с.

Як бачимо, в середньому економія часу менша, що пояснюється особливістю оброблення природної мови. Цього разу прискорення, досягнуте за допомогою *MapReduce* для *BiLSTM*, становило 2 (ситуація аналогічна попередній); для *RCNN* ~ 3.2 (для чотирьох вузлів результат збільшився до ~ 4.41).

Перейдемо до результатів точності класифікації для аналізу аудіо як сигналу (табл. 3).

Таблиця 3. *Точність класифікації для сигналу*

Інформація	**CNN**	**RNN**	**LSTM**	**BiLSTM**	**RCNN**
Вибори	0.89	0.91	0.93	0.95	0.97
Війна	0.91	0.89	0.92	0.97	0.97

Треба зауважити, що використання *MapReduce* не вплинуло на точність класифікації для обох наборів інформації, тому відповідні викладки було відкинуто.

З огляду на досягнутий результат *RCNN* гарантує найвищу точність класифікації (хоча різниця з *BiLSTM* не є суттєвою). Розгляд аудіо як тексту показав, що ситуація майже не змінюється (табл. 4).

Таблиця 4. *Точність класифікації для тексту*

Інформація	**CNN**	**RNN**	**LSTM**	**BiLSTM**	**RCNN**
Вибори	0.92	0.91	0.91	0.96	0.96
Війна	0.90	0.91	0.94	0.95	0.96

Останньою метрикою є розмір навчальної вибірки, необхідний для досягнення точності щонайменше 90%. Для цього проведено кілька ітерацій з поступовим збільшенням кількості записів від 5000 до 10000 (у разі аудіосигналів переважна більшість була результатом аугментації). Виявлено, що задана точність досягається за умов: 7000 записів для *CNN*; 7500 записів для *RNN*; 6600 записів для *LSTM*; 6100 записів для *BiLSTM*; 5800 записів для *RCNN*. Отже, економія мінімального допустимого обсягу інформації становить: 1700 для *RCNN*; 1400 для *BiLSTM*; 900 для *LSTM*; 500 для *CNN*; 0 для *RNN*.

Тепер можемо систематизувати добуті значення метрик та визначимо альтернативи, оптимальні за Парето, для оброблення аудіо як сигналу (табл. 5). Усі ненормалізовані значення були нормовані та округлені до сотих.

Таблиця 5. *Значення критеріїв, оптимальних за Парето, унаслідок аналізу аудіо*

Модель	**Збереження часу**	**Точність**	**Збереження обсягу інформації**
CNN	1.00	0.90	0.29
LSTM	0.59	0.93	0.53
BiLSTM	0.31	0.96	0.82
RCNN	0.00	0.97	1.00

На основі результатів можна обчислити значення лінійної адитивної згортки з ваговими коефіцієнтами. Для *CNN* маємо 0.849, для *LSTM* – 0.856, для *BiLSTM* – 0.881, а для *RCNN* – 0.876. Можна зауважити, що найбільш ефективною моделлю у виявленні факту фальсифікації для аудіо як сигналу є двоспрямована рекурентна нейромережа з підтримкою довгострокової та короткочасної пам'яті.

Однак різниця між *BiLSTM* та *RCNN* мало відчутна й може вважатися похибкою. Крім цього, суттєвий виграш у швидкості для *BiLSTM* частково нейтралізується за допомогою застосування технології *MapReduce*.

Перейдемо до систематизації результатів, досягнутих унаслідок класифікації аудіо як текстової інформації. Відповідні нормалізовані значення, оптимальні за Парето, наведені нижче (табл. 6).

На основі результатів можна обчислити значення лінійної адитивної згортки з ваговими коефіцієнтами. Для *CNN* маємо 0.771, для *LSTM* – 0.833, для *BiLSTM* – 0.918, а для *RCNN* – 0.917. Як і в попередньому разі, найефективнішою моделлю є *BiLSTM*, однак виграш у швидкодії зменшується з допомогою паралелізації.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Таблиця 6. *Значення критеріїв, оптимальних за Парето, унаслідок аналізу тексту*

Модель	Збереження часу	Точність	Збереження обсягу інформації	Контекст
CNN	1.00	0.91	0.29	0.60
LSTM	0.53	0.93	0.53	0.80
BiLSTM	0.27	0.96	0.82	1.00
RCNN	0.00	0.96	1.00	1.00

Беручи до уваги наведене вище, зазначимо, що найбільш ефективними моделями є *BiLSTM* та *RCNN*, а результати застосування технології *MapReduce* доводять доцільність її використання для класифікації фейкових аудіозаписів різного виду.

Висновки

Метою статті було розроблення ефективної моделі для визначення факту підробки звукової інформації з використанням технології *MapReduce*. Для цього проаналізовано особливості фальсифікації аудіоінформації у вигляді сигналу й тексту. Крім цього, досліджені сучасні наукові публікації, присвячені обраній темі, і низка експертних опитувань дали змогу сформувати набір алгоритмів для створення власної моделі визначення фейкових аудіо. Перша стадія цієї моделі передбачає:

– якщо аудіо – це текст: передоброблення інформації за допомогою *Google Speech to Text* та подальшу конвертацію тексту в числове подання, зважаючи на частотно-емоційній характеристики (знайдені за допомогою алгоритму BM-25) самого повідомлення та останніх перевірених новин, ступені надійності конвертації, вагу повідомлення;

– якщо аудіо – це сигнал: очищення сигналу від шуму та невокалізованих ділянок із подальшою аугментацією за допомогою векторної авторегресії, паралелізованої за допомогою *MapReduce* (результати експерименту дали змогу обгрунтувати вибір класичної векторної авторегресії).

Наступною стадією є застосування нейронної мережі. З огляду на проаналізовані дослідження обрано рекурентні та згорткові нейронні мережі, зокрема:

– класична згорткова нейромережа;

– класична рекурентна нейромережа;

– рекурентна нейромережа з довгостроковою пам’яттю;

– двоспрямована рекурентна нейромережа з довгостроковою пам’яттю;

– гібридна нейромережа, що поєднує декілька згорткових мереж із двоспрямованою рекурентною мережею з довгостроковою пам’яттю.

Щоб подолати проблему, пов’язну з часом навчання моделі, використано технологію *MapReduce*. Для визначення найбільш ефективної нейронної мережі та доцільності застосування запропонованого способу паралелізації сформовано набір критеріїв, що дав змогу використати принцип лінійної адитивної згортки з ваговими коефіцієнтами. На основі цих критеріїв, зазначених модифікацій та імплементації обраних моделей за допомогою бібліотек *Python 3* проведено серії експериментів з інформацією щодо виборчого процесу в Україні 2019 р. та повномасштабного вторгнення Російської Федерації на територію нашої країни.

У процесі експериментів виявлено, що двоспрямована рекурентна нейромережа з довгостроковою пам’яттю є найбільш ефективною, хоча вона й поступається у швидкості менш складним моделям. Водночас різниця в ефективності між нею та гібридною нейромережею не суттєва. З’ясовано, що виграш у економії часу передоброблення внаслідок застосування технології *MapReduce* може становити 4.3 – для тексту і 4 – для сигналу. Перевага, пов’язана з економією часу навчання нейромереж може досягати 4.71 – для тексту і 4.68 – для сигналу, нівелюючи розрив у ефективності між *BiLSTM* та *RCNN*.

Отже, використання побудованої моделі на основі *BiLSTM* (або *RCNN*) є високоефективним для визначення факту підробки аудіо як у вигляді тексту, так і сигналу. Упровадження передоброблення інформації та навчання нейромереж за допомогою *MapReduce* є доцільним. Відкритими залишаються проблеми розширення результатів на зображення та відеоматеріали, а також можливості застосування інших підходів для класифікації та аугментації інформації.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Список літератури

1. Anders M. Fake News Detection. European Data Protection Supervisor. URL: https://edps.europa.eu/press-publications/publications/techsonar/fake-news-detection_en (дата звернення: 27.05.2024).
2. Real-Time Advanced Computational Intelligence for Deep Fake Video Detection / N. Bansal та ін. *Applied Science*. 2023. Vol. 13 (5). 3095 p. DOI: 10.3390/app13053095
3. A Signal Detection Approach to Understanding the Identification of Fake News / C. Batailler та ін. *Perspectives on Psychological Science*. 2023. Vol. 17 (1). P. 78–98. DOI: 10.1177/1745691620986135
4. Reis J. Supervised Learning for Fake News Detection / J. C. S. Reis та ін. *IEEE Explore*. 2019. Vol. 34 (2). P. 76–81. DOI: 10.1109/MIS.2019.2899143
5. Giandomenico D. D. Fake news, social media and marketing: A systematic review / D. D. Giandomenico та ін. *Journal of Business Research*. 2021. Vol. 124. P. 329–341. DOI: 10.1016/j.jbusres.2020.11.037
6. Yuan L. Sustainable Development of Information Dissemination: A Review of Current Fake News Detection Research and Practice / L. Yuan та ін. *Systems*. 2023. Vol. 11 (9). 458 p. DOI: 10.3390/systems11090458
7. The impact of fake news on social media and its influence on health during the COVID-19 pandemic / Y. M. Rocha та ін. *Journal of Public Health.* 2023. Vol. 31. P. 1007–1016. DOI: 10.1007/s10389-021-01658-z
8. Alonso M. Dataset for multimodal fake news detection and verification tasks / A. Bondielli та ін. *Data in Brief*. 2024. Vol. 54. 110440 p. DOI: 10.1016/j.dib.2024.110440.
9. Tolosana R. Sentiment Analysis for Fake News Detection / Tolosana R. та ін. *Electronics.* 2021. Vol. 10 (11). 1348 p. DOI: 10.3390/electronics10111348
10. Afanasieva I. Deepfakes and beyond: A Survey of face manipulation and fake detection / Afanasieva I. та ін. *Information Fusion*. 2020. Vol. 64. P. 131–148. DOI: 10.1016/j.inffus.2020.06.014
11. Afanasieva Nataliia Application of Neural Networks to Identify of Fake News / N. Afanasieva та ін. *Computational Linguistics and Intelligent Systems*, Kharkiv, 20–21 квітня 2023 р. 2023. 3396 p. URL: https://ceur-ws.org/Vol-3396/paper28.pdf (дата звернення: 27.05.2024)
12. Bhatia T. Using transfer learning, spectrogram audio classification, and MIT app inventor to facilitate machine learning understanding. *Massachusetts Institute of Technology*. URL: https://dspace.mit.edu/handle/1721.1/127379 (дата звернення: 27.05.2024).
13. Breuer A., Eilat R., Weinsberg U. Friend or Faux: Graph-Based Early Detection of Fake Accounts on Social Networks. *Web Conference*, Taipei, 20–24 квіт. 2023. P. 1287–1297. DOI: 10.1145/3366423.3380204
14. Xia T., Chen X. A. Discrete Hidden Markov Model for SMS Spam Detection. *Applied Science*. 2020. Vol. 10 (14). 5011 p. DOI: 10.3390/app10145011
15. Najar F., Zamzami N., Bouguila S. Fake News Detection Using Bayesian Inference. *Information Reuse and Integration for Data Science*, Los Angeles, 30 черв. – 1 серп. 2019. P. 389–394. DOI: 10.1109/IRI.2019.00066
16. Montserrat D. Generative Autoregressive Ensembles for Satellite Imagery Manipulation Detection / D. M. Montserrat та ін. *Workshop on Information Forensics and Security*, New York, 6–11 груд. 2020. P. 1–6. DOI: 10.1109/WIFS49906.2020.9360909
17. Ning C., You F. Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. *Computers & Chemical Engineering*. 2019. Vol. 125. P. 434–448. DOI: 10.1016/j.compchemeng.2019.03.034
18. Sardar T. H., Ansari Z. An Analysis of Distributed Document Clustering Using MapReduce Based K-Means Algorithm. *Journal of The Institution of Engineers (India):Series B*. 2020. Vol. 101. P. 641–650. DOI: 10.1007/s40031-020-00485-2
19. Deng R., Duzhin, F. Topological Data Analysis Helps to Improve Accuracy of Deep Learning Models for Fake News Detection Trained on Very Small Training Sets. *Big Data and Cognitive Computing*. 2022. Vol. 6 (3). 74 p. DOI: 10.3390/bdcc6030074
20. Choudhary A., Arora A. Linguistic feature based learning model for fake news detection and classification. *Expert Systems with Applications*. 2021. Vol. 169. 114171 p. DOI: 10.1016/j.eswa.2020.114171
21. Khovrat A. Parallelization of the VAR Algorithm Family to Increase the Efficiency of Forecasting Market Indicators During Social Disaster / A. Khovrat та ін. *Information Technology and Implementation*, м. Kyiv, 30 лист. – 2 груд. 2022. P. 222–233. URL: https://ceur-ws.org/Vol-3347/Paper_19.pdf (дата звернення: 27.05.2024).

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

References

1. Anders M. "Fake News Detection. European Data Protection Supervisor", available at: https://edps.europa.eu/press-publications/publications/techsonar/fake-news-detection_en (last accessed 27.05.2024).
2. Bansal, N., Aljrees, T., Yadav, D. P., Singh, K. U., Kumar, A., Verma, G. K., Singh, T. (2023), "Real-Time Advanced Computational Intelligence for Deep Fake Video Detection", *Applied Science*, No. 13(5), 3095 p. DOI: 10.3390/app13053095
3. Batailler, C., Brannon, S. M., Teas, P. E., Gawronski, B. (2023), "A Signal Detection Approach to Understanding the Identification of Fake News", *Perspectives on Psychological Science*, No. 17(1), P. 78–98. DOI: 10.1177/1745691620986135
4. Reis, J. C. S., Correia, A., Murai, F., Veloso, A., Benevenuto, F. (2019), "Supervised Learning for Fake News Detection", *IEEE Intelligent Systems*, No. 34(2), P. 76–81. DOI: 10.1109/MIS.2019.2899143
5. Giandomenico, D. D., Sit, J., Ishizaka, A., Nunan, D. (2021), "Fake news, social media and marketing: A systematic review", *Journal of Business Research*, Vol. 124, P. 329–341. DOI: 10.1016/j.jbusres.2020.11.037
6. Yuan, L., Jiang, H., Shen, H., Shi, L., Cheng, N. (2023), "Sustainable Development of Information Dissemination: A Review of Current Fake News Detection Research and Practice", *Systems*, No. 11(9), 458 p. DOI: 10.3390/systems11090458
7. Rocha, Y. M., de Moura, G. A., Desiderio, G. A., de Oliveira, C. H., Lourenço, F. D., de Figueiredo Nicolete, L. D. (2023), "The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review", *Journal of Public Health*, Vol. 31, P. 1007–1016. DOI: 10.1007/s10389-021-01658-z
8. Alonso, M. A., Vilares, D., Gómez-Rodríguez, C., Vilares, J. (2021), "Sentiment Analysis for Fake News Detection", *Electronics*, No. 10(11), 1348 p. DOI: 10.3390/electronics10111348
9. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., Ortega-Garcia, J. (2020), "Deepfakes and beyond: A Survey of face manipulation and fake detection", *Information Fusion*, Vol. 64, P. 131–148. DOI: 10.1016/j.inffus.2020.06.014
10. Afanasieva, I., Golian, N., Golian, V., Khovrat, A., Onyshchenko, K. (2023), "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, available at: https://ceur-ws.org/Vol-3396/paper28.pdf (last accessed: 27.05.2023).
11. Afanasieva Nataliia Application of Neural Networks to Identify of Fake News (2023), / N. Afanasieva et al. *Computational Linguistics and Intelligent Systems*, Kharkiv, 20–21.04.2023. 3396 p. available at: https://ceur-ws.org/Vol-3396/paper28.pdf (last accessed: 27.05.2024)
12. Bhatia, N. (2020), "Using transfer learning, spectrogram audio classification, and MIT app inventor to facilitate machine learning understanding", *Massachusetts Institute of Technology*, available at: https://dspace.mit.edu/handle/1721.1/127379 (last accessed 27.05.2024)
13. Breuer, A., Eilat, R., Weinsberg, U. (2023), "Friend or Faux: Graph-Based Early Detection of Fake Accounts on Social Networks", *Web Conference, 20–24 April 2023, Taipei*, P. 1287–1297. DOI: 10.1145/3366423.3380204
14. Xia, T., Chen, X. A. (2020), "Discrete Hidden Markov Model for SMS Spam Detection", *Applied Science*, Vol. 10 (14), 5011 p. DOI: 10.3390/app10145011
15. Najar, F., Zamzami, N., Bouguila, S. (2019), "Fake News Detection Using Bayesian Inference", *Information Reuse and Integration for Data Science, 30 July – 1 August 2019, Los Angeles*, P. 389–394. DOI: 10.1109/IRI.2019.00066
16. Montserrat, D. M., Horváth, J., Yarlagadda, S. K., Zhu, F., Delp, E. J. (2020), "Generative Autoregressive Ensembles for Satellite Imagery Manipulation Detection". *Workshop on Information Forensics and Security (WIFS 2020): 12th IEEE International Workshop, New York, 6 December – 11 December 2020: IEEE*, P. 1–6. DOI: 10.1109/WIFS49906.2020.9360909
17. Ning, C., You, F. (2019), "Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming", *Computers & Chemical Engineering*, Vol. 125, P. 434–448. DOI: 10.1016/j.compchemeng.2019.03.034
18. Sardar, T. H., Ansari, Z. (2020), "An Analysis of Distributed Document Clustering Using MapReduce Based K-Means Algorithm", *Journal of The Institution of Engineers (India): Series B*, Vol. 101, P. 641–650. DOI: 10.1007/s40031-020-00485-2
19. Deng, R., Duzhin, F. (2022), "Topological Data Analysis Helps to Improve Accuracy of Deep Learning Models for Fake News Detection Trained on Very Small Training Sets", *Big Data and Cognitive Computing*, Vol. 6 (3), 74 p. DOI: 10.3390/bdcc6030074
20. Choudhary, A., Arora, A. (2021), "Linguistic feature based learning model for fake news detection and classification", *Expert Systems with Applications*, Vol. 169, Article 114171. DOI: 10.1016/j.eswa.2020.114171

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

21. Khovrat, A., Kobziev, V., Nazarov, A., Yakovlev, S. (2022), "Parallelization of the VAR Algorithm Family to Increase the Efficiency of Forecasting Market Indicators During Social Disaster". *Information Technology and Implementation (IT&I 2022): 9th Internaional Conference, Kyiv, 30 November – 2 December 2022: CEUR Workshop Proceedings.* No. 3347, P. 222–233, available at: https://ceur-ws.org/Vol-3347/Paper_19.pdf (last accessed: 27.05.2024).

Надійшла (Received) 30.05.2024

Відомості про авторів / About the Authors

Ховрат Артем Вячеславович – Харківський національний університет радіоелектроніки, аспірант, Харків, Україна; e-mail: artem.khovrat@nure.ua; ORCID ID: https://orcid.org/0000-0002-1753-8929

Khovrat Artem – Kharkiv National University of Radio Electronics, Postgraduate Student at the Department of Software Engineering, Kharkiv, Ukraine.

THE EFFICIENCY ASSESSMENT OF USING HYBRID NEURAL NETWORKS FOR THE DETECTION OF FORGED AUDIO DATA IN SOCIALLY ORIENTED SYSTEMS

The **subject** of the research is the problem of detecting falsified data, in particular in audio format, in socially oriented systems. The **goal** of the work is to develop an effective model based on recurrent and convolutional neural networks for determining the fact of forgery of sound data, using MapReduce technology for parallelization. The article addresses the following **tasks**: determining the features of audio in socially-oriented systems, conducting an analysis of algorithms for processing audio information both in the form of text and in the form of a signal, forming a list of target architectures of neural networks and revealing the features of their implementation, conducting an experimental test of effectiveness selected approaches. The following **methods** used are – analytical and inductive method for determining the target set of neural network architectures; expert assessment for the formation of the most influential efficiency factors; experimental, multi-criteria evaluation and statistical methods of data augmentation to determine the most effective model. The following **results** were obtained: an audio data reprocessing algorithm was developed for the possibility of using recurrent and convolutional networks. Several approaches to data classification using augmentation based on vector autoregression and MapReduce parallelization technology have been implemented. It was determined that the most effective model for the multi-criteria selection problem is a combination of a bidirectional recurrent neural network with support for short- and long-term memory with several convolutional networks. The advantages of using MapReduce technology to optimize training time and data processing are shown, and a set of open questions for further research and applied implementation is defined. **Conclusions**: he application of an analytical and inductive approach followed by experimental verification made it possible to develop an effective (with an accuracy of more than 96%) a mechanism for detecting fabricated data both in the form of a signal and in text form. The obtained result makes it possible to assert the feasibility of implementing the proposed approach, and, accordingly, makes it possible to reduce the influence of such information in socially oriented systems, especially during crisis events.

Keywords: signal augmentation; vector autoregression; classification; natural language processing; fake information.

Бібліографічні описи / Bibliographic descriptions

Ховрат А. В. Оцінювання ефективності використання гібридних нейронних мереж для виявлення сфальсифікованої аудіоінформації в соціально орієнтованих системах. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 166–181. DOI: https://doi.org/10.30837/2522-9818.2024.2.166

Khovrat, A. (2024), "The efficiency assessment of using hybrid neural networks for the detection of forged audio data in socially oriented systems", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 166–181. DOI: https://doi.org/10.30837/2522-9818.2024.2.166