

Продемонстрований метод має проблеми під час навчання, адже як в тестову вибірку, так і в навчальну будуть потрапляти одні і ті ж зображення, що може призводити до завчання моделлю результату замість знаходження закономірностей, зменшити негативний вплив можна змінюючи дублі: додаванням шумів, обертанням чи віддзеркаленням, зміною кольорів.

Крім продемонстрованого вище метода досягти зменшення негативного впливу дисбалансу класів можна введенням вагових коефіцієнтів для кожного класу, таким чином в обраному датасеті зросте штраф за нерозпізнаний корабель, а нагорода за розпізнану відсутність корабля впаде, що нівелює вплив дисбалансу на результат навчання моделі, однаке потребує додаткових обчислень.

Отже, продемонстровані методи боротьби з дисбалансом класів дозволяють уникати негативного впливу на якість передбачення навченої моделі, що було продемонстровано на прикладі обраного датасету.

Список використаних джерел

1. *Survey on deep learning with class imbalance* [Електронний ресурс]. Режим доступу до ресурсу: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0192-5>
2. *Japkowicz N. The class imbalance problem: Significance and strategies. In: In proceedings of the 2000 international conference on artificial intelligence (ICAI). 2000;111–7.*
3. *Ships in Satellite Imagery* [Електронний ресурс] // *Online Journal for Research and Education*. – 2021. – Режим доступу до ресурсу: <https://www.kaggle.com/datasets/rhammell/ships-in-satellite-imagery>

УДК 004.01

*Триконенко С.В., студент 2 курсу
магістратури
спеціальності 122 «Комп'ютерні науки»
Бабаков Р.М., доцент,
доцент кафедри інформаційних технологій*

РОЗРОБКА І ДОСЛІДЖЕННЯ МЕТОДІВ ПАРСИНГУ БІБЛІОГРАФІЧНИХ ДАНИХ

Донецький національний університет імені Василя Стуса, м. Вінниця

В Україні та в інших країнах світу, функціонують бібліотеки, які становлять основу культурної, наукової, інформаційної та освітньої інфраструктури. Серед них можна виділити публічні бібліотеки, які є центрами культурного життя суспільства. Головних завданням бібліотек є усунення

інформаційної нерівномірності шляхом гарантування рівних можливостей доступу до інформації для різних верств населення. Важливий зв'язок між бібліотеками та їх потенційними користувачами створюють інформаційні ресурси. Сучасна автоматизована бібліотечна інформація системи містять модуль «Електронний каталог», який замінює бібліотечні карткові каталоги та доступний через мережу Інтернет.

У науці залишається давно актуальна проблема інтеграції даних і розвивається одночасно з інформаційними технологіями та підвищенням вимог до якості даних. Дослідницька організація Online Computer Library Center для розширення доступу до інформації в усьому світі та скорочення витрати на інформацію займається розробкою методів парсингу бібліографічних даних.

Крім того, розроблено основи для глобальної інтеграції даних, зокрема: розробка моделі реляційної обробки неелементарних даних[1, 2]; методи інтеграції різнорідних баз даних; методи створення зведеної інформації ресурсів спільного доступу[3, 4]; концепція інтеграції інформаційних потоків[5]; методи інтеграції даних у відкритих системах[6]; аналіз та узагальнення методів і засобів парсингу даних [7].

Тому виникає проблема парсингу неоднорідних даних в цілому та зокрема бібліографічних даних. Загальні та спеціальні методи парсингу постійно розвиваються та вдосконалюються, що робить дану тему актуальною.

Мета роботи полягає в дослідженні особливостей парсингу бібліографічних даних шляхом створення інтелектуальної системи для узгодження та якісної інтеграції інформації.

Для досягнення поставленої мети необхідним вирішити наступні **завдання**:

- обробка та аналіз великого масиву бібліографічних даних, які є характерними для публічних бібліотек України;
- дослідження масиву даних з метою пошуку особливостей та закономірностей;
- аналіз головних проблем парсингу бібліографічних даних;
- розробка ефективних методів парсингу даних на основі використання особливостей та закономірностей;
- реалізації методів з метою їх використання для створення з електронного каталогу публічних бібліотек.

В даній роботі використано фреймворк Spark для обробки великих масивів даних, який здатен обробляти великі обсяги даних на кластерах комп'ютерів. Продемонстровано ряд альтернатив використання архітектури веб-додатків, а саме монолітний та мікросервісний, обрано мікросервісний підхід, який хоч і має ряд недоліків, але добре підходить в даному випадку, а можливість поділу системи на невеликі компоненти з подальшим розширенням його складових є чудовою альтернативою в даному випадку.

Також продемонстровано основні компоненти розробленої інтелектуальної системи. Сервіс прогнозування використовує описані підходи для прогнозування даних послідовності в майбутньому, а сервіс прийому даних,

використовуючи паралельну обробку, отримує та аналізує бібліографічні дані, після чого передає їх до відповідного сервісу обробки, де, в залежності від розміру даних та складності завдання, дані можуть бути оброблені одразу.

Отже, в даній роботі розглянуті основні принципи та характеристики бібліографічної інформації. Показано, що розробка інтелектуальної системи для таких задач є непростим завданням, це пов'язано з даними цих систем, які постійно оновлюються, а їх кількість надзвичайно велика, для реалізації та обробки даних з цих систем необхідно використовувати відповідні інструментальні засоби та архітектуру.

Продемонстровано основні алгоритми прогнозування послідовностей даних, а саме лінійна регресія та її вдосконалення, рекурсивні нейронні мережі, а також метод експоненціального спуску, який хоч і простий в реалізації, але його вдосконалення дають ефективні результати. Хоча кожен з цих методів має свої переваги та недоліки, їх робота продемонстрована на реальному застосуванні та вивчені їх основних принципів.

Список використаних джерел

1. Blei D. *Modeling annotated data* // *Proceedings of the 26th annual international acm sigir conference on research and development in informaion retrieval*. – New York, NY, USA: ACM, 2020. – P. 127-134.
2. Brin S. *The anatomy of a large-scale hypertextual Web search engine* // *Computer Networks and ISDN Systems*, 2019. – P. 107-117.
3. DBLP [Електронний ресурс] – Режим доступу до ресурсу: <https://dblp.org/> – (дата звернення: 15.09.2022).
4. Frost R. *Parser Combinators for Ambiguous Left-Recursive Grammars*. // *10th International Symposium on Practical Aspects of Declarative Languages (PADL), ACM-SIGPLAN, Volume 4902/2018, Pages:, January 2018, San Francisco*. – P. 167-181.
5. Grune D. *Parsing Techniques – A Practical Guide* / D. Grune, C. Jacobs. – Chichester: Originally published by Ellis Horwood, 2020. – 320 p.
6. Gundecha U. *Selenium WebDriver 3 Practical Guide: End-to-end automation testing for web and mobile browsers with Selenium WebDriver* / U. Gundecha, S. Avasarala. – 2020. – 280 p.
7. Гук-Сатайкін О. І., Бабаков Р. М., Засоби прикладного аналізу бібліографічних даних. *Прикладні інформаційні технології: матеріали II Всеукраїнської науково-практичної конференції для студентів, аспірантів та молодих вчених, м. Вінниця: ДонНУ, 29 квітня 2021.* – 28-29 с. URL: <https://jait.donnu.edu.ua/article/view/12328/>

УДК 004.021

Хмелівський Ю.С., студент СО
«Магістр» спеціальності 122
«Комп'ютерні науки»
Римар П.В., старший викладач
кафедри комп'ютерних наук та
інформаційних технологій

**ОПИС ПОКРАЩЕНОГО МЕТОДУ ОПТИМІЗАЦІЇ ТРАНСПОРТНОЇ
ЗАДАЧІ**