МЕТОДИ ДИФЕРЕНЦІЙНОЇ ПРИВАТНОСТІ В ГІБРИДНИХ СИСТЕМАХ РЕКОМЕНДАЦІЙ НА ОСНОВІ МАТРИЧНОЇ ФАКТОРИЗАЦІЇ

Лебьодкін Д. О.[1], Жульковський О.О.[1], Жульковська І.І.[2]
[1]*Дніпровський державний технічний університет, Кам'янське*
[2]*Університет митної справи та фінансів, Дніпро*

ВСТУП

Розвиток цифрових технологій призвів до того, що алгоритми машинного навчання (ML) та методи інтелектуального аналізу даних стали невід'ємною частиною сучасної інформаційної інфраструктури [1, 2]. Проте широке впровадження персоналізованих сервісів актуалізує питання захисту приватності. Дані, які використовуються для побудови моделей – від історії фінансових транзакцій до вподобань у контенті – можуть бути використані для несанкціонованого відновлення конфіденційної інформації про користувача. У зв'язку з цим виникає необхідність у дослідженні механізмів, які б дозволили поєднати високу аналітичну потужність моделей з гарантіями безпеки даних.

Актуальність і постановка проблеми. Сьогодні персоналізовані рекомендаційні системи є критично важливими для ефективної взаємодії з користувачем у мережі Інтернет. Вони дозволяють фільтрувати надлишок інформації, пропонуючи найбільш релевантний контент. Проте архітектура таких систем передбачає збір та обробку детальних профілів активності, що створює значні ризики приватності. Традиційні методи анонімізації даних часто виявляються неефективними проти сучасних атак, таких як атаки на інверсію моделі або висновок про належність.

Це зумовлює потребу в переході до методів, що мають під собою формальне математичне обґрунтування. Найбільш перспективним підходом на сьогодні є використання диференційної приватності (Differential Privacy, DP). Проблема полягає у пошуку такого балансу параметрів захисту, за якого похибка моделі залишається в межах допустимої норми для бізнес-логіки системи, а ризик витоку даних обмежується.

Мета дослідження. Оцінка ефективності та стійкості методів забезпечення конфіденційності в ML-моделях, аналіз компромісу між рівнем захисту даних і точністю передбачень, а також розробка практичних рекомендацій щодо інтеграції таких підходів у реальні сервіси.

Аналіз останніх досліджень і публікацій. Забезпечення приватності в сучасних ML-застосунках базується на математично формалізованих критеріях, серед яких центральне місце посідає DP. Вона гарантує, що зміна одного запису в навчальному наборі має обмежений вплив на поведінку моделі, що дозволяє встановлювати суворі математичні межі захисту. Питання витоку інформації через складні аналітичні запити та

вразливість моделей перед атаками на висновок про належність детально розглянуто в [3]. Дослідження підтверджують, що навіть агреговані дані можуть бути деанонімізовані без належного зашумлення, що вимагає впровадження формальних критеріїв приватності [4].

Особлива увага в науковій літературі приділяється інтеграції механізмів Лапласа та Ґауса безпосередньо в процеси оптимізації моделей. Проте, як зазначається у дослідженнях, присвячених великим рекомендаційним сервісам, таким як, наприклад, Netflix [5], впровадження захисту часто призводить до зниження точності прогнозів. Пошук балансу між безпекою та якістю персоналізації залишається актуальним викликом, що потребує вдосконалення методів обробки латентних факторів [6].

Короткий опис дослідження, його методів і засобів. Теоретичний фундамент роботи базується на формалізації взаємодій між користувачами та об'єктами через матрицю взаємодій R. Оскільки реальні системи працюють в умовах значної розрідженості даних, основним методом дослідження обрано матричну факторизацію на основі сингулярного розкладу (Singular Value Decomposition, SVD) [5].

В основі запропонованої архітектури лежить обробка вихідних даних через рівень DP. Загальна схема запропонованого підходу представлена у табл. 1, де показано процес проходження даних через модуль захисту перед етапом формування рекомендацій. На цьому рівні впроваджується контрольований шум згідно з параметром бюджету приватності ε. Зашумлені дані проектуються у латентний простір за допомогою SVD, що дозволяє виділити стабільні поведінкові патерни.

Таблиця 1

Послідовність трансформації даних у системі з DP

Етап	Процес	Результат
Вхідні дані	Збір сирих оцінок, кліків та уподобань користувачів	Матриця взаємодії R
Захист (DP)	Внесення контрольованого шуму Лапласа/Ґауса згідно з бюджетом ε	Зашумлені (спотворені) дані
Факторизація	SVD для виділення латентних векторів p_u та q_i	Приватно-збережені латентні фактори
Генерація	Прогнозування рейтингів через скалярний добуток векторів	Персоналізований список релевантних об'єктів

Представлений у таблиці підхід дозволяє математично знизити розмірність обчислень, замінюючи аналіз великої розрідженої матриці компактним представленням у вигляді латентних векторів p_u та q_i. При цьому прогнозний рейтинг обчислюється як їх скалярний добуток.

Це дозволяє моделі узагальнювати інформацію навіть за відсутності прямих збігів в історії оцінок. Водночас латентні фактори акумулюють найбільш інформативну частину персонального профілю користувача, що робить їх особливо чутливими з точки зору приватності.

У колаборативних моделях фундаментальну роль відіграє оцінка подібності між профілями. Найбільш поширеною метрикою є косинусна подібність. Такий підхід дозволяє оцінювати саме спрямованість уподобань, а не абсолютну величину активності, що важливо для систем із нерівномірним розподілом рейтингів. На відміну від колаборативних методів, контентно-орієнтовані підходи базуються на аналізі характеристик об'єктів за допомогою моделі TF-IDF (Term Frequency – Inverse Document Frequency), яка дозволяє формалізувати текстові описи у числові вектори [6].

Застосування TF-IDF у поєднанні з матричною факторизацією дозволяє спроектувати гібридну модель, здатну обробляти як історію взаємодій, так і атрибути контенту. Для забезпечення формальних гарантій захисту в роботі використано принцип ε-DP.

Практична реалізація полягає у внесенні контрольованого шуму безпосередньо в процес ітераційного навчання. Це дозволяє нівелювати вплив одиничних записів на фінальні параметри латентних векторів, забезпечуючи стійкість системи до деанонімізації. Ефективність даного підходу підтверджується результатами експериментальних досліджень.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Експериментальне дослідження на метаданих платформи RAWG.io було спрямоване на аналіз впливу бюджету приватності ε на якість функціонування гібридної системи. Основна увага приділялася виявленню точки балансу, де гарантії конфіденційності не призводять до критичної втрати релевантності рекомендацій. Для оцінки точності було використано метрику Precision@10, а для вимірювання похибки прогнозування – середньоквадратичну помилку (Root Mean Square Error, RMSE).

Колаборативні методи на основі SVD-факторизації продемонстрували вищу стійкість до шуму, тоді як контентно-орієнтовані алгоритми виявилися більш чутливими до внесення диференційного збурення. Це пояснюється здатністю SVD зберігати домінуючі структурні закономірності у латентному просторі, ефективно відфільтровуючи стохастичні спотворення.

Інтегрований аналіз результатів, представлених у табл. 2, демонструє чітко виражений компроміс (trade-off) між рівнем захисту та корисністю моделі.

За умов суворих обмежень приватності ($\varepsilon = 0.1$) система демонструє максимальну деградацію метрик: значення Precision@10 знижується до 0.46, а RMSE сягає 1.00. Такий стан відповідає інтенсивному внесенню шуму, який фактично перекриває корисний сигнал латентних факторів.

Таблиця 2

Залежність Precision@10 та RMSE від ε

Epsilon (ε)	Precision@10	RMSE
0.1	0.46	1.00
0.5	0.51	0.92
1.0	0.55	0.86
2.0	0.65	0.68
5.0	0.81	0.56

Зі збільшенням бюджету приватності показники якості монотонно покращуються. Оптимальний режим експлуатації моделі спостерігається в інтервалі $\varepsilon = 1.0–2.0$. У цьому діапазоні точність Precision@10 зростає від 0.55 до 0.65, а RMSE знижується з 0.86 до 0.68. Саме ці значення можна вважати найбільш вигідним компромісом для комерційних систем, де необхідно забезпечити високу релевантність при дотриманні математичних стандартів захисту персональних даних.

Подальше послаблення обмежень ($\varepsilon \to 5.0$) призводить до поступової стабілізації показників. При $\varepsilon = 5.0$ точність сягає 0.81, а RMSE становить 0.56, проте ризики деанонімізації при таких значеннях зростають, а приріст точності стає менш вираженим порівняно зі зростанням ризиків. Це підтверджує гіпотезу про те, що використання латентних представлень, отриманих за допомогою SVD-факторизації, дозволяє ефективно згладжувати випадкові спотворення, зберігаючи працездатність моделі навіть під значним тиском захисних механізмів.

Отримані результати [7] підтверджують можливість практичної інтеграції механізмів диференційної приватності у рекомендаційні системи, наприклад [8].

ВИСНОВКИ

Проведено комплексне дослідження методів забезпечення конфіденційності в інтелектуальних системах на основі принципів диференційної приватності та матричної факторизації.

Встановлено, що застосування процедури зниження розмірності на основі сингулярного розкладу сприяє стабілізації моделі в умовах зашумленості. Завдяки апроксимації низького рангу, алгоритм дозволяє виділити домінуючі поведінкові патерни (латентні фактори), відсіюючи при цьому стохастичні збурення, що вносяться механізмами диференційної приватності. Це дозволяє зберегти топологічну структуру рекомендаційного поля навіть при низьких значеннях бюджету приватності.

При $\varepsilon = 2.0$ середньоквадратична похибка знижується до 0.68, а при $\varepsilon = 1.0$ становить 0.86, що є прийнятним рівнем для більшості комерційних сценаріїв. Оптимальним з точки зору балансу є значення $\varepsilon = 2.0$, де одночасно досягається Precision@10 = 0.65 та RMSE = 0.68.

Встановлено, що колаборативна фільтрація на основі SVD демонструє суттєво вищу стійкість до шумового впливу порівняно з контентно-орієнтованими підходами. Це зумовлено тим, що латентні фактори акумулюють стабільні поведінкові патерни, які є більш резистентними до збурень Лапласа, ніж прямі оцінки у матриці взаємодій. Інтеграція контентного компонента у гібридній моделі підвищує загальне покриття системи, але не її шумову стійкість.

Перспективи подальших досліджень полягають у розробці адаптивних алгоритмів додавання шуму, які б враховували щільність даних у конкретних кластерах користувачів, що дозволить ще більше знизити похибку без втрати гарантій конфіденційності.

ДЖЕРЕЛА

1. Шумейко О.О., Сотник В.С., Жульковська І.І., Жульковський О.О. Використання методів Data Mining для обробки мовної інформації. *Математичне моделювання.* 2021. № 2 (45). С. 48–57. URL: https://doi.org/10.31319/2519-8106.2(45)2021.246944

2. Vokhmianin H., Zhulkovska I., Zhulkovskyi O., Ulianovska Yu., Mala Yu. Forecasting demand for products using neural models and time series. Mathematical Modeling. 2024. Vol. 50 (1). pp. 19–31. URL: https://doi.org/10.31319/2519-8106.1(50)2024.304779

3. Shokri R., Stronati M., Song C., Shmatikov V. Membership inference attacks against machine learning models. *Proc. 2017 IEEE Symposium on Security and Privacy (SP)*. 2017. P. 3–18. URL: https://doi.org/10.1109/SP.2017.41

4. Dwork C., Roth A. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science.* 2014. 9(3–4). P. 211–407. URL: https://doi.org/10.1561/0400000042

5. Gomez-Uribe C. A., Hunt N. The Netflix recommender system: algorithms, business value, and innovation. *ACM Transactions on Management Information Systems*. 2016. 6(4). Article 13. URL: https://doi.org/10.1145/2843948

6. Zhang S., Yao L., Sun A., Tay Y. Deep learning based recommender system: a survey and new perspectives. *ACM Computing Surveys*. 2019. 52(1). Article 5. URL: https://doi.org/10.1145/3285029

7. Zhulkovska I., Zhulkovskyi O., Yakovenko V., Rudianova T., Mala Yu., Lebedkin D. Approaches to Ensuring Data Privacy in Machine Learning Models. *Mathematical Modeling*. 2026. Vol. 54. No 1. pp. 49–57. URL: https://doi.org/10.31319/2519-8106.1(54)2026.352417

8. Zhulkovskii O., Panteikov S., Zhulkovskaya I. Information-modeling forecasting system for thermal mode of top converter lance. *Steel in Translation.* 2022. Vol. 52. No. 5. P. 495–502. URL: https://doi.org/10.3103/S0967091222050138