Технічні науки

УДК 004.4

Романенко Лев Анатолійович

бакалавр програмної інженерії

Національного технічного університету України

«Київський політехнічний інститут імені Ігоря Сікорського»

Романенко Лев Анатольевич

бакалавр программной инженерии

Национального технического университета Украины

«Киевский политехнический институт имени Игоря Сикорского»

Romanenko Lev

Bachelor of software engineering

The National Technical University of Ukraine

«Igor Sikorsky Kyiv Polytechnic Institute»

ІНТЕГРУВАННЯ АЛГОРИТМУ РОЗПОДІЛЕНОГО МАШИННОГО НАВЧАННЯ І МЕХАНІЗМУ ДИФЕРЕНЦІАЦІЇ КОНФІДЕНЦІЙНОСТІ В СИСТЕМУ КРАУДСЕНСІНГУ

ИНТЕГРИРОВАНИЯ АЛГОРИТМА РАСПРЕДЕЛЕННОГО МАШИННОГО ОБУЧЕНИЯ И МЕХАНИЗМА ДИФФЕРЕНЦИАЦИИ КОНФИДЕНЦИАЛЬНОСТИ В СИСТЕМУ КРАУДСЕНСИНГА

INTEGRATING DISTRIBUTED MACHINE LEARNING ALGORITHM AND DIFFERENTIAL PRIVACY MECHANISM INTO THE CROWDSENSING SYSTEM

Анотація. *Портативні інтелектуальні пристрої такі як мобільні телефони зі вбудованими сенсорами і доступом в інтернет стали основою всіх інтелектуальних особистих гаджетів. Велика кількість пристроїв мають можливість колективно збирати і виконувати обробку даних в безпрецедентних масштабах. В даній роботі представлене програмне*

забезпечення що зберігає конфіденційність машинного навчання для групи смартфонів що дозволяє вирішити широкий спектр проблем пов'язаних з машинним навчанням групи пристроїв з диференціальними умовами конфіденційності. Система надає можливість навчати класифікатори чи програми прогнозування онлайн, на даних з краудсенсінгу, приватно та з мінімальними обчислювальними затратами на пристроях та серверах.

Ключові слова: *краудсенсінг, диференціальна приватність, машинне навчання.*

Аннотация. *Портативные интеллектуальные устройства такие как мобильные телефоны со встроенными сенсорами и доступом в интернет стали основой всех интеллектуальных личных гаджетов. Большое количество устройств имеют возможность коллективно собирать и выполнять обработку данных в беспрецедентных масштабах. В данной работе представлено программное обеспечение сохраняющая конфиденциальность машинного обучения для группы смартфонов, которое позволяет решить широкий спектр проблем, связанных с машинным обучением группы устройств с дифференциальными условиями конфиденциальности. Система предоставляет возможность обучать классификаторы или программы прогнозирования онлайн, на данных из краудсенсингу, приватно и с минимальными вычислительными затратами на устройствах и серверах.*

Ключевые слова: *краудсенсинг, дифференциальная приватность, машинное обучение.*

Summary. *Portable intelligent devices such as mobile phones with built-in sensors and Internet access have become the basis of all intelligent personal gadgets. A large number of devices have the ability to collectively collect and perform data processing on an unprecedented scale. In this paper, a software that preserves the confidentiality of machine learning for a group of*

smartphones is presented, which allows solving a wide range of problems related to machine learning of a group of devices with differential privacy conditions. The system provides the ability to teach classifiers or forecasting programs online, on cursor data, privately and with minimal computing costs on devices and servers.

Key words: *crowdsensing, differential privacy, machine learning.*

Краудсенсінг

Розумні пристрої стають все більш поширеними в повсякденному житті. Ці пристрої характеризуються вбудованими датчиками (наприклад, акселерометри, камери, мікрофони), обчислювальною здатністю і підключення до Інтернету за допомогою бездротового зв'язку або стільникових мереж. До них відносяться стаціонарні пристрої, наприклад прилади розумного дому або мобільні пристрої, такі як смартфони. Все більше і більше пристроїв поєднуються між собою, це явище називають «інтернетом речей». Взаємозв'язок надає можливості для груп розумних пристроїв колективно обмінюватися і обробляти дані на безпрецедентних масштабах. Запропоновано різні застосування краудсенсінгу, включаючи моніторинг особистого здоров'я / фітнесу, екологічні зондування та моніторинг дорожніх умов. Список стрімко продовжується розширюватися.

Краудсенсінг використовується в основному для збору та аналізу сукупних даних з групи учасників. Однак, можна виконати більш складні та корисні завдання поза розрахунком сукупної статистики, за допомогою алгоритмів машинного навчання на даних краудсенсінгу. Приклади таких завдань включають:

- вивчення оптимальних параметрів кімнатних температур для смарт-термостатів;
- пропонування найкращого маршруту для водіння;

– розпізнавання звуків притаманних конкретній події з мікрофона.

Алгоритми і типи даних для даних завдань є різні, але всі вони можуть бути навчені стандартно без контролю навчання або під контролем: враховуючи сенсорні дані (час, розташування, рух, заміри датчиків навколишнього середовища тощо), тренується алгоритм або модель, яка може точно передбачити змінну вподобань (установка температури, поточна активність користувача, трафік, аудіо події тощо). Умовно, краудсенсінг і машинне навчання виконується у вигляді двох окремих процесів: збирання і відправлення даних до центрального агрегатора та процеси аналізу або навчання що виконуються на сервері.

Конфіденційність

Конфіденційність є важливою проблемою для додатків, що займаються збором даних. Забезпечуючи конфіденційність учасників, краудсенсінг-система може об'єднати більшу кількість потенційних учасників, що підвищує корисність такої системи. Однак багато систем краудсенсінгу описаних в джерелах не використовують жодного механізму збереження конфіденційності. Протягом останнього десятиліття популярності набула диференційна конфіденційність як формальний показник ризику втрати конфіденційності даних. В загальному, диференціальна конфіденційність вимірює, як в значній мірі результат процедури змінюється ймовірнісно присутністю або відсутністю якого небуть об'єкта в оригіналі даних [1, с. 83]. Ця міра забезпечує верхню границю втрати конфіденційності незалежно від яких небуть попередніх значень що може мати зловмисник [2, с. 23]. В той час як диференціальна конфіденційність була оприлюднена в наукових виданнях і використана в машинному навчанні у неї не було широкого застосування в системах краудсенсінгу. В цій статті інтегрується диференціально-конфіденційні механізми в краудсенсінг системах, які мають змогу забезпечити надійний захист від різних способів атак.

Схема роботи системи

Система складається з сервера і декількох інтелектуальних пристроїв (смартфонів), здатних до збору сенсорних даних, чисельних обчислень та комунікацій з сервером за допомогою глобальної мережі інтернет. Мета роботи - навчання класифікатора або прогнозу вподобань на даних зібраних за допомогою декількох пристроїв. Широкий діапазон класифікаторів або предикторів можна навчати шляхом загального методу статистичного навчання пов'язаного з даним завданням - мінімізації емпіричного ризику. Нехай $x \epsilon R^D$ буде вектором обробки даних з сенсорів, (аудіо, відео, акселерометр і т.д.), а y - цільова змінна, ціль якої зробити передбачення з x, наприклад діяльність користувача. Для регресії, у може бути дійсним числом, а для класифікації y – дискретна мітка $y \in \{1, \dots, C\}$ з C класами. Дані визначаємо як N пар (ознака вектора, цільова змінна), що генеруються i.i.d. від невідомого розподілу усіма пристроями, що беруть участь, до наступного:

$$D = \{(x_1, y_1), \dots, (x_N, y_N)\} \quad (1)$$

Припустимо, що ми використовуємо класифікатор / предиктор $h\,(x;\, w)$ з змінним параметром вектора w, і функцією втрат $l\,(y, h\,(x;\, w))$ для вимірювання продуктивність класифікатора / предиктора по відношенню до істинної цілі y. Широкий спектр алгоритмів навчання може бути представлений h і l, наприклад, регресія, логістична регресія, і машина опорних векторів. Якщо там є M розумних пристроїв, ми знаходимо оптимальні параметри w класифікатора/предиктора шляхом мінімізації емпіричного ризику усіх M пристроїв:

$$\mathcal{R}(w) = \sum_{m=1}^{M} \frac{1}{|\mathcal{D}_m|} \sum_{(x,y)\in\mathcal{D}_m} l(h(x; w), y) + \frac{\lambda}{2}\|w\|^2 \quad (2)$$

де D_m - набір вибірок, створених тільки з пристрою m, і $\frac{\lambda}{2}||\omega||^2$ є виразом регуляризації. Ця функція ризику (2) може бути мінімізована за допомогою багатьох методів оптимізації. У цій роботі використовуємо

стохастичний (суб) градієнтний спуск (СГС) [3, с. 34], який є одним з найпростіших методів оптимізації і також підходить для широкомасштабного навчання. СГС мінімізує ризик оновленням w послідовно

$$\omega(t+1) \leftarrow \prod_{w}[\omega(t) - \eta(t)g(t)], \quad (3)$$

де (t) - швидкість навчання, а $g\ (t)$ – градієнт функція втрати

$$g = \nabla_{\omega} l(h(x;\omega), y), \quad (4)$$

оцінюється зразком $(x;\ y)$ і поточним параметром $w\ (t)$. Будемо вважати, що параметр домену W є d-мірним кулька деякого великого радіуса R, а проекція $\Pi_w = \min\left(1, \frac{R}{||w||}\right) w$. За замовчуванням ми використовуємо швидкість навчання

$$\eta^{(t)}\frac{c}{\sqrt{t}}, \quad (5)$$

де c - постійний гіперпараметр. При обчисленні градієнтів, використовується "мініатюр" з b вибірок для обчислення усереднений градієнт

$$\tilde{g} = \frac{1}{b}\sum_i \nabla_{\omega} l(h(x_i;\omega), y_i), \quad (6)$$

який відіграє важливу роль у компромісі продуктивність-конфіденційність і масштабованості. У ПЗ ризик мінімізації за допомогою СГС здійснюється шляхом розподілу основного навантаження (= обчислення усереднених градієнтів) до M пристроїв. Важливо те, що кожен пристрій генерує дані та обчислює градієнти використовуючи власні дані. Робочий процес описаний на

Механізм конфіденційності

У системах краудсенсінгу особисті дані користувачів можуть втратити конфіденційність багатьма шляхами. Наприклад системні адміністратори або аналітики можуть навмисно "зливати" інформації або ж витік може статися при публікації аналітики даних. Також існують більш складні шляхи втрати конфіденційності даних, наприклад

перехоплення пристроями що маскуються під штатні, за допомогою хакерських даних що в процесі роботи системи зберігаються на сервері або підслуховуванням між пристроями і серверами.

Замість того, щоб розробляти окремий механізм захисту для кожного типу атаки, в статті розробляється єдиний локальний метод який буде реалізований на кожному пристрої для всіх типів атак. Даний алгоритм обробляє певним чином дані перед тим як вони покинуть пристрій.

Локальний механізм враховує, що зловмисник потенційно може отримати доступ до всіх переданих даних між пристроями і сервером, що також включає інші типи атак. Це через те що дані:

1) видно зловмисним пристроям;
2) зберігаються на сервері;
3) випромінені учасниками системи, можуть бути перехвачені між пристроями і сервером.

Нехай локальна ϵ-диференціальна приватність буде кількісною мірою конфіденційності. Формально, «рандомізований» алгоритм, який приймає дані $\mathcal{D}$ в якості вхідних і вихідних даних f називається ϵ-диференціальною приватністю, якщо

$$\frac{P(f(\mathcal{D})\in S)}{P(f(\mathcal{D}`)\in S)} \leq e^{\epsilon} \quad (7)$$

для всіх вимірюваних $S \subset T$ діапазону виходу і для всіх наборів даних $\mathcal{D}$ і $\mathcal{D}`$ відрізняються одним елементом [1]. Тобто, навіть якщо зловмисник має всі дані $\mathcal{D}$, за винятком одного елемента, це не дозволяє аналізувати одиницю даних з виводу алгоритму f. Менше ϵ робить такий висновок більш складним, і тому робить алгоритм більш зберігаючим приватність. Коли алгоритм виводить реальний вектор $f \in R^D$, його глобальна чутливість може бути визначена шляхом

$$S(f) = \max_{\mathcal{D},\mathcal{D}`} \|f(\mathcal{D}) - f(\mathcal{D}`)\|_1 \quad (8)$$

де $||\cdot||_1$ - норма L_1. Основний результат з визначення диференціальної приватності є те, що вектор-функція f з чутливість $S(f)$ може бути зроблена ϵ-диференційно-приватною шляхом додавання незалежного шумового вектора Лапласа z, де

$$P(z) \propto e^{-\frac{\epsilon}{S(f)}\|z\|_1} \quad (9)$$

У даній системі розглядається ϵ-диференційна конфіденційність будь-якого одного (особливість, мітка) - прикладу, виявленого за допомогою комунікації між усіма пристроями і сервером, які є градієнтами $\tilde{g}$, числа зразків n_s, число неправильно класифікованих зразків n_e, і мітки, обчислюються як n_y^k [2]. Кількість необхідного шуму залежить від вибору функцій втрати.

Prediction	$\arg\max_k w_k' x$
Risk	$\mathcal{R}(w) = \frac{1}{N}\sum_i[-w'_{y_i}x_i + \log\sum_l e^{w_l' x_i}] + \frac{\lambda}{2}\sum_k \|w_k\|^2$
Gradient	$\nabla_{w_k}\mathcal{R} = \frac{1}{N}\sum_i x_i[-I[y_i = k] + P(y = k\|x_i)] + \lambda w_k$

Рис. 1. Мультикласова логістична регресія

Це значення обчислюється для багатокласової логістичної регресії (рис 1), але вона може бути обчислена аналогічно для інших функцій втрат. Додаючи елементарно незалежний шум Лапласа z до усереднених градієнтів $\tilde{g}$

$$\hat{g} = \frac{1}{b}\sum_i g_i + z,\ P(z) \propto e^{-\frac{\epsilon g^b}{4}|z|}, \quad (10)$$

існує така гарантія конфіденційності:

Теорема 1 (усереднене градієнтне збурення). Передача $\tilde{g}$ за формулою. (10) ϵ_g-диференційно приватний.

Для захисту даних, n_e і n_y^k, додаємо дискретний шум Лапласа наступним чином:

$$\widehat{n_e} = n_e + z, P(z) \propto e^{-\frac{\epsilon_e}{2}|z|}, \quad (11)$$

$$\widehat{n_y^k} = n_y^k + z, P(z) \propto e^{-\frac{\epsilon_{y^k}}{2}|z|}, \quad (12)$$

Де $z = 0, \pm 1, \pm 2, \ldots$. Ці механізми мають такі гарантії конфіденційності:

Теорема 2 Передача n_eі n_y^kза формулами. (11) і (12) - ϵ_eі - ϵ_{y^k} – диференційно приватні, відповідно.

Практично, системний адміністратор вибирає ϵ в залежності від бажаного рівня конфіденційності для зібраних даних. Невеликої $\epsilon(\to 0)$ можна використовувати для даних, які користувачі вважають дуже приватними, наприклад поточне місце знаходження, а також великий $\epsilon(\to \infty)$ для менш приватних даних, таких як температура навколишнього середовища.

Висновки. Дана система інтегрує алгоритми розподіленого навчання і механізми диференціації конфіденційності в систему краудсенсінгу. Загалом, робота має такі наукові внески:

- Створено загальну основу для машинного навчання з допомогою смарт-пристроїв від crowdsensing даних з багатьма потенційними додатками.
- Реалізовано різні гарантії конфіденційності, які забезпечують надійний механізм конфіденційності від різних типів атак в краудсенсінгу
- .Дана робота є першим ПЗ, що інтегрує збір даних, навчання і диференційовані приватні механізми для краудсенсінгу.
- Проведений аналіз ПЗ для демонстрації того що обчислювальні затрати на реалізацію конфіденційності можна мінімізувати. Також затрати на обчислення і комунікацію є не великими що дозволяє розгортати систему у великих масштабах.

Література

1. Дворко К. «Диференціальна конфіденційність», в автоматах, мовами і програмуванні. – Springer, 2006. – С. 1-12.
2. Вапник В. Природа статистичної теорії навчання. – Springer, 2000.
3. Роббінс Г., Монро С. "Метод стохастичної апроксимації", Аннали математичної статистики, 1951. – С. 400–407.
4. Лейн Н. Д., Мілуццо Е., Лу Х., Піблз Д., Чоудхурі Т., Кемпбелл А. Т. Дослідження сприйняття мобільного телефону, вересень 2010. - С. 140–150.