Vulnerability Analysis of Privacy-Preserving Algorithms in Multidimensional Medical and Nutritional Databases

Matyoqubov Og‘abek Bahrombek o‘g‘li
Urganch State University
Yusupov Bahodir Karomatovich
Institute of Information and Communication Technologies and Military Communications

Abstract

This paper analyzes the vulnerabilities of privacy-preserving algorithms applied to multidimensional medical and nutritional databases. Syntactic anonymization models (k-anonymity, l-diversity, t-closeness), differential privacy and federated learning are compared. It is shown that each of them is exposed to specific attacks: homogeneity and semantic similarity attacks, privacy budget exhaustion under composition, and gradient inversion. The need for adaptive, dynamic privacy models is substantiated.

Keywords

privacy, medical data, nutritional data, k-anonymity, differential privacy, federated learning, re-identification.

Ko‘p o‘lchamli tibbiy va ovqatlanish ma’lumotlar bazalarida maxfiylikni saqlovchi algoritmlarning zaifliklari tahlili

Matyoqubov Og‘abek Bahrombek o‘g‘li
Urganch davlat universiteti
Yusupov Bahodir Karomatovich
Axborot-kommunikatsiya texnologiyalari va harbiy aloqa instituti

Annotatsiya. Maqolada ko‘p o‘lchamli tibbiy va ovqatlanish ma’lumotlar bazalarida qo‘llaniladigan maxfiylikni saqlovchi algoritmlarning zaifliklari tahlil qilinadi. Sintaktik anonimlashtirish modellari (k-anonimlik, l-xilma-xillik, t-yaqinlik),

differensial maxfiylik va federativ o‘qitish qiyosiy o‘rganilgan. Ularning har biri o‘ziga xos hujumlarga: gomogenlik va semantik o‘xshashlik hujumlariga, kompozitsiya natijasida maxfiylik byudjetining sarflanishiga hamda gradient inversiyasiga zaif ekani ko‘rsatilgan. Adaptiv va dinamik maxfiylik modellarini ishlab chiqish zarurligi asoslangan.

Kalit so‘zlar: maxfiylik, tibbiy ma’lumotlar, ovqatlanish ma’lumotlari, k-anonimlik, differensial maxfiylik, federativ o‘qitish, qayta identifikatsiyalash.

Kirish

Raqamli tibbiyot, mobil salomatlik (mHealth) va tibbiy buyumlar interneti (IoMT) tizimlari foydalanuvchilarning salomatlik ko‘rsatkichlari hamda ovqatlanish ratsioniga oid ma’lumotlarni uzluksiz to‘plamoqda. Ovqatlanish odatlari diabet, gipertoniya va allergiya kabi holatlarni bilvosita ochib beradi, shuning uchun bunday ma’lumotlar nozik toifaga kiradi.

Faqat to‘g‘ridan-to‘g‘ri identifikatorlarni olib tashlash yetarli emas: L. Sweeney tadqiqotida AQSh aholisining 87 foizi faqat jinsi, tug‘ilgan sanasi va pochta indeksi bo‘yicha yagona aniqlanishi ko‘rsatilgan [1].

Foydalanuvchi profili va hujum turlari

Foydalanuvchi profili $P_u = \langle DI_u, QI_u, SA_m(t), SA_n(t) \rangle$ ko‘rinishida ifodalanadi: bunda DI – to‘g‘ridan-to‘g‘ri identifikatorlar, QI – kva-identifikatorlar, $SA_m(t)$ va $SA_n(t)$ – vaqt bo‘yicha o‘zgaruvchi tibbiy va ovqatlanish atributlari. Karbogidrat iste’moli va glyukoza, natriy va qon bosimi o‘rtasidagi korrelyatsiya yashirin tashxisni ochib beradi. Asosiy tahdidlar: qayta identifikatsiyalash, atribut va a’zolik xulosasi.

Sintaktik anonimlashtirish modellarining zaifliklari

k-anonimlik modeli ekvivalentlik sinfi E uchun $|E| \geq k$ shartini talab qiladi va qayta identifikatsiyalash ehtimolini $1/k$ bilan chegaralaydi [1]. Biroq sinfdagi barcha yozuvlarda nozik atribut bir xil bo‘lsa (masalan, barchasida diabet tashxisi), tajovuzkor

aniq yozuvni ajratmasa ham kasallikni bilib oladi (gomogenlik hujumi). *l*-xilma-xillik modeli [2] bu kamchilikni bartaraf etadi, ammo semantik o‘xshash qiymatlar (oshqozon yarasi, gastrit, duodenit) oldida himoyasiz qoladi. *t*-yaqinlik modeli [3] sinf ichidagi taqsimot global taqsimotdan *t* dan ortiq farq qilmasligini talab qiladi, lekin atributlar soni ko‘payganda ma’lumotlarning tibbiy tahlil uchun yaroqliligini keskin pasaytiradi. Qiyosiy tahlil 1-jadvalda berilgan.

1-jadval. Sintaktik anonimlashtirish modellarining qiyosiy tahlili

Model	**Matematik sharti**	**Himoya qila olmaydigan zaiflik**	**Tibbiy va ovqatlanish ma’lumotlaridagi kamchiligi**
k-anonimlik	$\|E\| \geq k$	Gomogenlik va fon bilimlari hujumi	Bir xil nozik atribut kasallikni oshkor qiladi
l-xilma-xillik	$H(E) \geq \log l$	Semantik o‘xshashlik hujumi	O‘xshash tashxislar sirni fosh etadi
t-yaqinlik	$D[P, Q] \leq t$	Vaqtli qatorlar korrelyatsiyasi hujumi	Real bog‘liqlikni buzadi

Differensial maxfiylik va federativ o‘qitish zaifliklari

Tasodifiy *M* algoritm bir yozuvi bilan farq qiluvchi har qanday D_1 va D_2 to‘plamlar uchun $P[M(D_1) \in S] \leq e^{\varepsilon} \cdot P[M(D_2) \in S] + \delta$ tengsizlikni qanoatlantirsa, (ε, δ)-differensial maxfiylikni ta’minlaydi [4]. Usulning ikki zaifligi bor. Birinchidan, bir foydalanuvchining uzluksiz kiritiladigan ma’lumotlari bo‘yicha ketma-ket so‘rovlarda maxfiylik byudjeti yig‘iladi: $\varepsilon_{jami} = \Sigma\varepsilon_i$, natijada himoya darajasi pasayadi. Ikkinchidan, kichik ε qiymatida kiritilgan shovqin klinik jihatdan muhim qiymatlarni buzadi.

Federativ o‘qitishda ma’lumot qurilmada qoladi, serverga esa faqat model yangilanishlari uzatiladi: $W_{t+1} = W_t + \Sigma\,(n_k/N)\Delta w_k$ [5]. Biroq uzatilgan gradientlar orqali dastlabki o‘quv namunasini qayta tiklash mumkinligi isbotlangan [6]. DP-SGD usuli [7]

bu xavfni kamaytiradi, ammo model aniqligini pasaytiradi. Gomomorfik shifrlash esa mobil qurilmalar uchun hisoblash jihatidan og‘ir.

Natijalar 1-rasmda umumlashtirilgan.



1-rasm. Himoya algoritmlari va ularning asosiy zaifliklari sxemasi

Xulosa

Hech bir algoritm ko‘p o‘lchamli tibbiy va ovqatlanish ma’lumotlarida maxfiylik, foydalilik va hisoblash samaradorligini bir vaqtda to‘liq ta’minlay olmaydi. Sintaktik modellar gomogenlik va semantik hujumlarga, differensial maxfiylik byudjetning kompozitsion sarfiga, federativ o‘qitish esa gradient inversiyasiga zaif. Shu sababli maxfiylik byudjetini atributlar kritikligiga qarab adaptiv taqsimlovchi kompleks modellarni ishlab chiqish zarur.

Foydalanilgan adabiyotlar ro‘yxati

[1] Sweeney L. k-anonymity: a model for protecting privacy // International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. – 2002. – Vol. 10, No. 5. – P. 557–570.

[2] Machanavajjhala A., Kifer D., Gehrke J., Venkitasubramaniam M. l-diversity: privacy beyond k-anonymity // ACM Transactions on Knowledge Discovery from Data. – 2007. – Vol. 1, No. 1. – Article 3.

[3] Li N., Li T., Venkatasubramanian S. t-closeness: privacy beyond k-anonymity and l-diversity // Proc. IEEE 23rd International Conference on Data Engineering (ICDE). – 2007. – P. 106–115.

[4] Dwork C., Roth A. The algorithmic foundations of differential privacy // Foundations and Trends in Theoretical Computer Science. – 2014. – Vol. 9, No. 3–4. – P. 211–407.

[5] McMahan H.B., Moore E., Ramage D., Hampson S., Agüera y Arcas B. Communication-efficient learning of deep networks from decentralized data // Proc. AISTATS. – 2017. – P. 1273–1282.

[6] Zhu L., Liu Z., Han S. Deep leakage from gradients // Advances in Neural Information Processing Systems 32 (NeurIPS). – 2019. – P. 14774–14784.

[7] Abadi M. et al. Deep learning with differential privacy // Proc. ACM CCS. – 2016. – P. 308–318.