Konfidensiallığı qorumaqla fərdi məlumatların intellektual analizi üçün Deep Learning metodları

Yadigar İmamverdiyev
AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
yadigar@iit.science.az

***Xülasə*— Bu məqalədə fərdi məlumatların intellectual analizi zamanı konfidensiallığın qorunması metodlarının, o cümlədən Deep Learning metodlarının icmalı verilir. Son onilliklərdə fərdi məlumatlar kibercinayətkarların əsas hədəflərindən birinə çevrilmişdir, konfidensiallığın qorunması tədbirlərinə baxmayaraq, verilənlərin intellektual analizi metodları fərdlər barədə sensitiv informasiya əldə etməyə imkan verir. Bu təhdidin qarşısını almaq üçün ənənəvi intellektual analiz metodlarının konfidensial fərdi məlumatlarla işləyən çoxsaylı modifikasiyaları təklif edilmişdir. Bu işdə fərdi məlumatların konfidensiallığının qorunması üçün dərin təlim metodlarının tətbiqi vəziyyəti analiz edilir, dərin təlim metodlarının diferensial gizlilik, çoxtərəfli təhlükəsiz hesablamalar və homomorf şifrləmə metodları ilə birləşdirilməsində qarşıya çıxan çətinliklər və onların həlli yolları araşdırılır.**

Açar sözlər— fərdi məlumatlar, gizlilik, gizlilik təhdidləri; PPDM, dərin təlim, Deep Learning

I. GİRİŞ

Big data şəxsi həyatın gizliliyinə, vətəndaş azadlıqlarının pozulmasına potensial təhdidlər yaradır, dövlət və korporativ nəzarət imkanlarını artırır. Big Data analitikasından istifadə etməklə şəxs barəsində gizli məlumatlar əldə etmək mümkündür. Analitika üçün verilənlərin anonimləşdirilməsi istifadəçi verilənlərinin konfidensiallığını qorumaq üçün kifayət deyil. Buna görə də, fərdi məlumatların intellektual analizi zamanı konfidensiallığın pozulması hallarının qarşısını almaq üçün müvafiq yanaşma, metod və texnologiyaların işlənilməsi vacibdir [1].

PPDM (Privacy-preserving data mining – konfidensiallığı qorumaqla verilənlərin intellektual analizi) yanaşmasının məqsədi data mining (DM) və ya maşın təlimi metodlarının tətbiqi ilə alınmış gizli informasiyaya icazə olmayan istifadəçilərin girişinin qarşısını almaqdır. Tədqiqatçılar konfidensiallığı qorumaq üçün PPDM-də data mining və maşın təlimi alqoritmlərində bir çox üsullardan istifadə edirlər [2, 3].

Son onilliklərdə fərdi məlumatlar kibercinayətkarların əsas hədəflərindən birinə çevrilmişdir, kütləvi informasiya vasitələri tez-tez böyük həcmli fərdi məlumatların oğurlanması halları barədə xəbərlər dərc edir. Texnologiyaların inkişafı fərdi məlumatların təhlükəsizliyinin təmin edilməsinin qanunvericilik bazasına yenidən baxılmasını zəruri edir. Avropa İttifaqının 2018-ci ilin mayından qüvvəyə minmiş fərdi məlumatların konfidensiallığının təmin edilməsi üzrə yeni qanunu GDPR (General Data Protection Regulation) bu sahədə ciddi tədbirlər nəzərdə tutur. Fərdi məlumatların emalı ilə məşğul olan təşkilatlar onların konfidensiallığının təmin edilməsi üzrə texnologiyaları praktikada tətbiq etməyə başlayırlar. Məsələn, Apple (iOS / macOS üçün), Google (RAPPOR), LinkedIn (Salary), Microsoft (Windows telemetriyası üçün) diferensial gizlilik sistemlərini praktikada tətbiq edir [4].

Son dövrlər dərin təlim (ing. Deep Learning) verilənlərin intellektual analizinin əsas metodlarından birinə çevrilmişdir və bir sıra praktiki məsələlərin həlli üçün geniş tətbiq edilir [5, 6]. Hazırda dərin təlim dərin neyron şəbəkələri ilə eyniləşdirilir. Dərin neyron şəbəkəsi (Deep Neural Network, DNN) – bir neçə gizli neyron layı olan neyron şəbəkəsidir. Dərin təlim xidmət kimi (Deep Learning as a Service, DLaaS) bulud platforması dərin neyron şəbəkələrinin geniş praktiki tətbiqlərinə yol açır. Google, Amazon və IBM öz bulud xidmətlərində DLaaS platformalarını işə salıblar [7]. Müxtəlif DLaaS istifadə etməklə müxtəlif gizli verilənlərini bulud serverinə göndərir. Server DNN tətbiqini yerinə yetirir və nəticələri müştəriyə geri göndərir. Aydındır ki, müştərinin verilənlərinin konfidensiallığı qorunmasa, DLaaS istifadəsi potensial gizlilik problemləri yarada bilər. Dürüst, lakin maraqlı bulud serveri müştərinin konfidensial verilənlərindən sensitiv məlumatlar toplaya bilər. Bu qəbildən olan problemləri araşdırmaq üçün fərdi məlumatların konfidensiallığının təmin edilməsi üçün bir sıra tədqiqatlar aparılmışdır.

Bu işin məqsədi verilənlərin intellektual analizi zamanı fərdi məlumatların konfidensiallığının qorunması üçün dərin təlim metodlarının tətbiqi vəziyyətini analiz etməkdir. Qeyd edək ki, [8, 9]-da bu məsələyə baxılır. Bu işdə həmin məqalələrdə əhatə edilməyən mənbələr tədqiqata cəlb edilmişdir və mövcud metodların dərin təlim metodlarına əsaslanan fərqli təsnifatı təklif edilmişdir. Təqdim olunan məqalədə əvvəlcə klassik PPDM metodları analiz edilir, daha sonra baxılan məsələ üçün dərin təlim metodlarının tətbiqi məsələləri araşdırılır.

II. FƏRDİ VƏ KOLLEKTİV KONFİDENSİALLIQ

Fərdi konfidensiallığın qorunması

Verilənlərin konfidensiallığının təmin edilməsinin əsas məqsədi – fərdi məlumatların qorunmasıdır. Ümumiyyətlə, məlumat o zaman fərdi hesab olunur ki, onu birbaşa və ya

DOI: 10.25045/NCInfoSec.2019.17

dolayısı ilə fiziki şəxslə əlaqələndirmək mümkün olsun. Beləliklə, fərdi məlumatlar intellektual analiz edilərkən fiziki şəxslə əlaqəli atributların qiymətləri məlum olur ki, onların konfidensiallığı qorunmalıdırlar.

Kollektiv konfidensiallığın qorunması

Fərdi məlumatların qorunması yetərli olmaya bilər. Bəzən qrupun fəaliyyətini əks etdirən gizli biliklərin əldə edilməsinin qarşısını almaq tələb edilə bilər. Gizli biliklərin qorunmasını *konfidensiallığın kollektiv qorunması* adlandırırlar. Burada məqsəd statistik verilənlər bazasında olduğu kimidir, yəni təhlükəsizliyə nəzarət mexanizmləri qruplar barədə informasiyanın aqreqasiyasını təmin edir, eyni zamanda fiziki şəxslər haqqında informasiyanın qarşısını alırlar. Lakin statistik verilənlər bazalarından fərqli olaraq, konfidensiallığın kollektiv qorunmasının digər məqsədi bütün statistika təhriflərinin minimallaşdırılması deyil, strateji qərarlar üçün birinci dərəcəli əhəmiyyəti olan strateji modelin qorunmasıdır. Başqa sözlə, burada məqsəd təkcə fərdi məlumatların deyil, açıqlanmaması tələb edilən bəzi modelləri və tendensiyaları da qorumaqdır.

III. VERİLƏNLƏRİN GİZLİLİYİNƏ TƏHDİDLƏR

Bu bölmədə fərdlərin gizli və/və ya sensitiv (həssas) məlumatlarının açıqlanması ilə əlaqəli təhdidlərə baxılır.

Konfidensiallıq təhdidləri atributların üç müxtəlif növü ilə əlaqəlidir [10]: birbaşa identifikatorlar, kvazi-identifikatorlar və sensitiv atributlar. Birbaşa identifikatorlar fərdləri dəqiq (aşkar) re-identifikasiya edən atributlardır, məsələn, ad, poçt ünvanı, telefon nömrəsi, milli ID-lər, e-poçt ünvanı və s. Kvazi-identifikatorlar kombinasiya edildikdə fərdi identifikasiya etməyə imkan verirlər (məsələn, cins, doğum tarixi və poçt indeksi, demoqrafik məlumatlar və diaqnoz kodları). Sensitiv atributlar fərdlərin (pasiyentləri) əlaqələndirilmək istəmədikləri atributlardır. Bu atributlara misallar spesifik diaqnoz kodları (məsələn, psixi xəstəlik, SPİD, xərçəng) və genom məlumatlarıdır (DNT).

Yuxarıdakı atributların növləri əsasında konfidensiallıq təhdidlərinin aşağıdakı növlərinə baxmaq olar:

Şəxsiyyətin açıqlanması (və ya re-identifikasiya): bu nəşr olunan verilənlər bazalarında ən məşhur təhdid hesab olunur. Bədniyyətli müəyyən şəxsi nəşr olunmuş bazadakı yazı ilə əlaqələndirdikdə baş verir.

Üzvlüyün açıqlanması [11]: Bu təhdid bədniyyətlinin müəyyən bir şəxsə aid yazının nəşr olunmuş verilənlər bazasında olması haqqında yüksək ehtimalla nəticə çıxardıqda baş verir. Məsələn, yalnız SPİD-pozitiv pasiyentlər haqqında məlumatlar olan verilənlər bazasına baxaq. Bir şəxsə aid yazının həmin bazada olması faktından onun SPİD-pozitiv olması nəticəsini çıxarmağa imkan verir və bu gizliliyə təhdiddir.

Atributun açıqlanması (və ya sensitiv informasiyanın açıqlanması): Bu təhdid şəxsi onun sensitiv məlumatları ilə əlaqələndirdikdə baş verir. Belə məlumatlar, məsələn, pasiyentin DNT məlumatları və ya şəxsin sensitiv məlumatlarının interval qiymətləri ola bilər (məsələn, müalicə xərclərinin təxmini kəmiyyəti).

Qeyd edək ki, ədəbiyyatda PPDM və PPDP (Privacy-Preserving Data Publishing) metodlarına qarşı bir çox hücumlar təklif edilmişdir [10, 12, 13]. Burada GAN (Generative Adversarial Network) hücumu [14] və MIA (Model Inversion Attack) hücumu [15] barəsində qısa məlumat verilir.

GAN hücumunda diskriminativ şəbəkə (D) generativ şəbəkəyə (G) qarşı 0-cəmli oyun oynayır. Oyun o zaman dayanır ki, D daha real və saxta şəkilləri fərqləndirə bilmir, yəni G D-ni aldatmaqda uğur qazanır [14].

MIA hücumunda maşın təlimi modelinə giriş icazəsi var, müəyyən fon məlumatı və onun sinfinin etibarlıq qiyməti (və ya sadəcə nişanı) verilib: verilmiş sinif üçün sinif ehtimalını maksimallaşdıran giriş tapılır. Məsələn, şəkli klassifikasiya edən model şəxsin adına əsasən şəxsin sifət şəklini generasiya edir **[**15**]**.

IV. GİZLİLİYİN QORUNMASI METODLARININ TƏSNİFATI

Fərdi məlumatların təhlükəsizliyinə cavabdeh təşkilatlar adətən de-identifikasya üsullarından, o cümlədən, anonimlik (*anonymization*), təxəllüs (*pseudonymization*), şifrləmə (*encryption*), açar-kodlaşdırma (*key-coding*) və s. istifadə edirlər. Anonimlik ad, ünvan və sosial təhlükəsizlik nömrələrini silməklə konfidensiallığı təmin edirsə, təxəllüs bu informasiyanı ləqəb və süni identifikasiya ilə əvəz edir. Açarla kodlaşdırma fərdi məlumatları kodlaşdırır və onların dekodlaşdırılması üçün açar yaradır [1,16].

PPDM və PPDP üçün çoxsaylı yanaşmalar mövcuddur (Cədvəl 1).

CƏDVƏL 1. GİZLİLİYİN QORUNMASI METODLARI

Yanaşma	Metodlar
Ümumi halda verilənlərin konfidensiallığının qorunması	Perturbasiya, randomizasiya, blok-lama, qarşılıqlı dəyişmə, seçmə, şifrləmə
Verilənlərin intellektual analizi zamanı konfidensiallığın qorunması	Assosiativ qaydaların çıxarılması, klassifikasiya, klasterizasiya
Verilənlərin nəşri zamanı konfidensiallığın qorunması	*k*-anonimlik, *l*-müxtəliflik, *m*-invariantlıq, *t*-yaxınlıq

PPDM metodlarını aşağıdakı meyarlar üzrə klassifikasiya etmək olar [17]:

- verilənlərin paylanması;
- verilənlərin modifikasiyası;
- verilənlərin intellektual analizi alqoritmləri;
- DM çıxışının gizlədilməsi.

Verilənlərin paylanması. Bəzi yanaşmalar mərkəzləşdirilmiş verilənlər üçün təklif edilib, digərləri paylanmış verilənlər üçün nəzərdə tutulub.

Bir neçə yerə paylanma halında problemi müəyyən etmək üçün verilənlərin necə paylanması mühüm rol oynayır. Müxtəlif bölünmələr fərqli problemlər yaradır və bu verilənlərin intellektual analizi zamanı konfidensiallığı qorumaq üçün müxtəlif alqoritmlərlə nəticələnir. Verilənlərin paylanması *üfüqi* və ya *şaquli* ola bilər.

Şaquli bölünmə. Verilənlərin şaquli bölünməsində nəzərdə tutulur ki, müxtəlif tərəflər eyni subyektlər çoxluğu barədə müxtəlif atributlar üzrə informasiya toplayırlar. Məsələn, banklar maliyyə tranzaksiyalarının məlumatlarını, vergi orqanları isə vergi məlumatlarını toplayırlar.

Üfüqi bölünmə. Üfüqi bölünmədə müxtəlif tərəflər müxtəlif subyektlər barədə eyni verilənlər çoxluğunu toplayırlar. Məsələn, müxtəlif supermarketlər ərzaq alınması barədə verilənləri toplayırlar.

Verilənlərin modifikasiyası: Bu metod konfidensiallığı qorumaq üçün relizdən öncə verilənlər bazası yazılarını modifikasiya edir. Verilənlərin modifikasiyası metodlarına perturbasiya, bloklama, aqreqasiya, qarşılıqlı dəyişmə, seçmə (populyasiyanın müəyyən hissəsi üçün verilənlər seçilir) daxildir [18]. Bu metodların hamısında atributların qiymətləri dəyişdirilir. Verilənləri üçüncü tərəf data maynerə verməzdən öncə verilənlərdə dəyişikliklər etmək lazımdır ki, mayner sensitiv informasiya barədə heç bir şey əldə edə bilməsin.

Verilənlərin perturbasiyası üçün iki əsas yanaşma var [18]: ehtimal paylanması yanaşması və verilənlərin qiymətlərinin təhrifi yanaşması. Ehtimal paylanması yanaşmasında verilənlər həmin paylanmadan olan digər nümunələrlə əvəzlənir. Verilənlərin qiymətlərinin təhrifi yanaşmasında verilənlər additiv küy, multiplikativ küy və ya digər randomizasiya proseduru ilə perturbasiya edilir. Adətən, küy qiymətini generasiya etmək üçün Qauss paylanması istifadə edilir. Küylərin korrelyasiyası ilkin verilənlərə nə qədər oxşardırsa, konfidensiallıq bir o qədər çox qorunur. Randomizasiya edilmiş verilənlərdən gizli verilənlərin çıxarılması imkanını qiymətləndirmək üçün PCA (Principal Component Analysis) və BE (Bayes Estimate) üsulları geniş öyrənilmişdir [19].

Verilənlərin qarşılıqlı dəyişdirilməsi (data swapping) metodunda verilənlər bazası sensitiv atributların qiymətini yazılar arasında qarşılıqlı dəyişməklə transformasiya edilir [20].

k-anonimlik modelində verilənlərin ümumiləşdirilməsi və verilənlərin silinməsi metodları tətbiq edilir və verilənlər toplusu yalnız o zaman nəşr olunur ki, həmin topluda olan hər bir şəxs üçün məlumatlar ən azı $(k-1)$ digər şəxsin məlumatlarından fərqləndirilə bilməsin [21].

Verilənlərin nəşri metodları, yəni *k*-anonimlik, *l*-müxtəliflik, *t*-yaxınlıq, *m*-invariantlıq dəqiqliyi və verilənlərin faydalılığını böyük dərəcədə azaldır.

Verilənlərin intellektual analizi alqoritmləri. Xüsusi olaraq PPDM üçün klassifikasiya, klasterləşdirmə, assosiativ qaydalar, Bayes şəbəkələri və s. alqoritmlərinin yeni variantları təklif edilmişdir [2]. Konfidensiallığı qorumaqla öyrənmə və konfidensiallığı qorumaqla klassifikasiya bu tədqiqatlarda iki əsas istiqamətdir. Konfidensiallığı qorumaqla klassifikasiya metodunda data mayner verilənləri klassifikasiya etməyi öyrənir, lakin bu zaman sensitiv atributları qorunur, mayner sensitiv verilənlər haqqında məlumat əldə edə bilmir.

DM çıxışının konfidensiallığı. DM alqoritmlərinin çıxışları bədniyyətli üçün çox informativ ola bilər. Məsələn, tutaq ki, assosiativ qayda mayninqi alqoritmində aşağıdakı qayda yüksək etibarlılıqla generasiya edilib:

$$(\text{Age} = 26, \text{ZIP code} = 10562) \Longrightarrow \text{HIV}.$$

Bu qaydanın aşkarlanması fərd barəsində olduqca gizli məlumatın açıqlanması ilə nəticələnir [2]. DM çıxışının konfidensiallığını qorumaq üçün geniş yayılmış üsullar aşağıdakılardır:

a) Assosiativ qaydaların gizlədilməsi – sensitiv qaydaların aşkarlanmasının qarşısını alır; bu məqsədlə verilənlərin modifikasiyası və bloklanması metodları istifadə edilir.

b) Klassifikatorun effektivliyinin aşağı salınması – verilənlərin sanitarizasiyası vasitəsilə klassifikatorun dəqiqliyi aşağı salınır ki, sensitiv məlumatların əldə edilməsi imkanı azalsın [22].

c) sorğuların auditi və məntiqi çıxarışlara nəzarət – aqreqativ verilənlər sorğuları ardıcıllığından informasiyanın açıqlanmasının qarşısını almaq məqsədi güdür. Sorğuların auditi zamanı sorğular ardıcıllığından bir və ya bir neçə sorğudan imtina edilir. İmtina edilən sorğular elə seçilir ki, verilənlərin sensitivliyi qorunur. Məntiqi çıxarışlara nəzarətdə sorğunun nəticəsi və ya verilənlər perturbasiya edilir [23].

Verilənlərin selektiv modifikasiyası və ya sanitarizasiyası NP-çətin məsələdir, bu səbəbdən mürəkkəblik probleminin həlli üçün evristik metodlar istifadə edilə bilər.

PPDM alqoritmlərinin işlənməsində vacib aspekt müvafiq qiymətləndirmə meyarlarının müəyyən edilməsi və uyğun qiymətləndirmə parametrlərinin (göstəricilərinin) işlənməsidir.

Konfidensiallığı qoruyan bütün alqoritmlər arasından bütün mümkün kriteriyalar üzrə digər alqoritmlərdən üstün olan alqoritm yoxdur. Alqoritm yalnız konkret kriteriya, məsələn, məhsuldarlıq üzrə digərindən üstün ola bilər. Buna görə istifadəçilərə onları maraqlandıran konkret parametrlərə görə ən yararlı alqoritmi seçməyə imkan verən metrikalar çoxluğu təqdim etmək lazımdır.

Konfidensiallığın qorunması alqoritmlərinin keyfiyyətini qiymətləndirmək üçün konfidensiallığı qoruma səviyyəsi, alqoritmin məhsuldarlığı, alqoritmin dəqiqliyi (gizlətmə metodu tətbiq edildikdən sonra PPDM nəticələrinin keyfiyyəti), verilənlərin faydalılığı (verilənlərin keyfiyyəti) və mürəkkəblik (konfidensiallığı qoruyan alqoritmin nəzərdə tutulan bütün resurslar baxımından yaxşı nəticələrlə yerinə yetirilməsi qabiliyyəti) kimi meyarlar istifadə edilə bilər [2].

V. FƏRDİ MƏLUMATLARIN GİZLİLİYİNİN QORUNMASINA ƏSAS YANAŞMALAR

PPDM-də verilənlərin konfidensiallığını təmin etmək üçün iki yanaşma istifadə edilir: verilənlərin təhrif edilməsi metodları (ing. data perturbation) və təhlükəsiz hesablama metodları (şifrləmə metodları). Bu iki yanaşmaya əsaslanan çoxsaylı məqalələr nəşr olunmuşdur.

Diferensial gizlilik

Diferensial gizlilik (Differential Privacy, DP) metodu ilk dəfə 2005-ci ildə təklif edilib [24]. Bu metod fərdi məlumatlara müraciətləri alqoritm və interfeys (etibarlı kurator kimi çıxış edir) vasitəsilə idarə edir. Tədqiqatçı verilənləri analiz etmək üçün kuratora sorğu göndərir və kurator da təsadüfi küy əlavə etməklə verilənlərin konfidensiallığını qorumağa və eyni zamanda sorğulara düzgün cavab verməyə çalışır. DP şirkətlərə istifadəçilərin vərdişləri barədə aqreqasiya olunmuş məlumatları toplamağa və bu zaman hər bir istifadəçinin konfidensiallığını qorumağa imkan verir.

Laplas mexanizmi. DP metodlarının bir çoxu verilənlərə idarə edilən küy əlavə edir. Laplas mexanizmi Laplas küyü əlavə edir (yəni Laplas paylanmasına tabe olan küy). DP mexanizmi yaratmağa imkan verən digər küy formaları, məsələn, Qauss küyü də əlavə etmək olar.

Diferensial gizliliyin tərifində iki D və D' verilənlər çoxluğuna baxılır [25]. D-də verilmiş fərdin məlumatları var və bu məlumatlar D'-də silinib (və ya hər hansı əlaqəsi olmayan məlumatlarla əvəzlənib). Tələb olunur ki, prosesin sonunda D və D'-ni bir-birindən fərqləndirmək mümkün olmasın. Fərqləndirməmə ε parametrindən istifadə etməklə ölçülür. ε parametri D və D'-dən alınmış nəticələrin paylanmasının nə dərəcədə yaxın olmasını ölçür. ε kiçik olduqda gizliliyin təhlükəsizliyi səviyyəsi yüksək olur.

Tərif. Əgər ən çoxu bir sıra ilə fərqlənən D və D' verilənlər çoxluqları və $S \subseteq Range(K)$ üçün aşağıdakı şərt ödənirsə, onda randomizasiya edilmiş K alqoritmi ε-diferensial gizlilik verir:

$$\Pr[K(D) \in S] \leq \exp(\varepsilon) \times \Pr[K(D') \in S].$$

DP alqoritmlərində iki kəmiyyətə baxılmalıdır:

Epsilon (ε): verilənlərdə diferensial dəyişiklik zamanı (bir sıra əlavə edildikdə və ya silindikdə) gizlilik itkisini ölçür, bu kəmiyyət nə qədər kiçikdirsə, gizlilik bir o qədər çox qorunur.

Dəqiqlik: DP alqoritminin çıxışının təmiz çıxışa yaxınlığı.

Laplas mexanizmindən fərqli mexanizmlər üçün daha bir kəmiyyət: *delta* (δ) istifadə edilir.

DP geniş spektrdə *gizlilik hücumlarına* (o cümlədən *diferensiallama hücumu*, *əlaqələndirmə hücumu* və *rekonstruksiya hücumlarına*) qarşı gizliliyin qorunmasının riyazi isbatlanan zəmanətini təmin edir [26].

Verilənlərin müxtəlif növləri üzərində analitik məsələlər üçün DP alqoritmlərinin işlənməsi sahəsində bir çox tədqiqat mövcuddur [26]. DP alqoritmlərinə misal olaraq RNM (Report Noisy Max) [26] və PATE (*Private Aggregation of Teacher Ensembles*) [27] alqoritmlərini göstərmək olar.

Təhlükəsiz çoxtərəfli hesablamalar

Paylanmış verilənlər üzərində PPDM kontekstində kriptoqrafiyaya əsaslanan üsullar aşağıdakı məzmunlu məsələləri həll etmək üçün işlənmişdi: iki və daha çox tərəf onların gizli verilənləri əsasında hesablamalar aparmaq istəyir. Hesablamaları elə aparmaq lazımdır ki, heç bir tərəf öz daxil etdiyindən və nəticədən başqa heç bir şey bilməsin. Bu məsələ təhlükəsiz çoxtərəfli hesablamalar (Secure Multiparty Computation, SMC) məsələsi adlanır. Təhlükəsiz ikitərəfli hesablama konsepsiyası formal olaraq 1986-cı ildə [28]-də daxil edilmiş, 1987-ci ildə isə [29]-da çoxtərəfli varianta genişləndirilmişdi.

Ümumiyyətlə, SMC kriptoqrafiyanın bir qoludur və paylanmış tapşırıqların təhlükəsiz şəkildə reallaşdırılmasını öyrənir. Burada təhlükəsizlik dedikdə konfidensiallığın təmin edilməsi, hesablamanın bədniyyətli hücumlardan qorunması və s. nəzərdə tutula bilər. SMC-nin məqsədi kollaborativ hesablama məsələsini həll etməkdir, yəni bir-birinə etibar etməyən qrupda hər hansı etibarlı üçüncü tərəf olmadan gizliliyi qorumaqla istifadəçilərin konfidensial verilənlərindən asılı açıq funksiyanın qiymətini hesablamaqdır. Aydındır ki, bu məsələ, mahiyyətcə, həm də PPDM məsələsidir və PPDM məsələsi SMC məsələsinin xüsusi halıdır.

Formal olaraq, tutaq ki, təhlükəsiz hesablamada $p_1, p_2, \ldots, p_n$ tərəfləri iştirak edir. Onların gizli verilənləri uyğun olaraq, $d_1, d_2, \ldots, d_n$-dir. İştirakçılar öz verilənlərinin konfidensiallığını pozmadan $f(d_1, d_2, \ldots, d_n)$ funksiyasının qiymətini hesablamaq istəyirlər, burada f bütün iştirakçılara məlum olan açıq funksiyadır.

İştirakçılar bir-birinə və ya kommunikasiya etdikləri kanallara etibar etməyə bilərlər. İştirakçıların arasında yarı-dürüstlərin olmasına yol verilir, yəni onlar protokola əməl edirlər, lakin istənilən aralıq verilənlərdən əlavə informasiya almağa cəhd edirlər.

Lindell və Pinkas ilk dəfə SMC üsulunu üfüqi bölünmüş verilənlər üzərində ID3 alqoritmi istifadə edilməklə klassifikasiya üçün təklif etmişdilər [30]. Ondan sonra şaquli bölünmüş verilənlərdə klassifikasiya və üfüqi bölünmüş verilənlərin klasterləşdirilməsi üçün bir sıra SMC alqoritmləri təklif edilmişdir [2].

Homomorf şifrləmə

Texnoloji yanaşmalardan biri də bu sahədə yeni kriptoqrafik alqoritmlərin işlənməsidir. Klassik şifrləmə üsulları artıq yetərli deyil və şifrlənmiş verilənlər üzərində hesablamalar aparmağa imkan verən üsulların işlənməsi tələb edilir:

Homomorf şifrləmə sistemi şifrlənmiş məlumatlar üzərində məlumatları deşifrləmədən riyazi əməliyyatlar (məsələn, toplama, çıxma, birləşmə, kəsişmə) aparmağa imkan verir. Məlumdur ki, RSA kriptosistemi məxfi açarı bilmədən şifrlənmiş verilənlər üzərində vurma əməlini yerinə yetirməyə imkan verir. Belə sistem *multiplikativ homomorf şifrləmə*

adlanır [31]. Oxşar olaraq, əgər məxfi açarı bilmədən toplama əməli dəstəklənirsə, *additiv homomorf şifrləmə* adlanır. *Tam homomorf şifrləmə* (Fully Homomorphic Encryption, FHE) məxfi açarı bilmədən istənilən hesablamanı dəstəkləyir, yəni istənilən $\odot$ əməli və iki m_1 və m_2 açıqmətnləri üçün

$$Enc(m_1) \odot Enc(m_2) = Enc(m_1 \odot m_2).$$

Belə sistemlərin qurulması onilliklər boyu kriptoqrafiyada açıq məsələ olmuşdur. Nəhayət, ilk dəfə 2009-cu ildə tam homomorf şifrləmə sistemi IBM şirkətinin kriptoqrafı C.Gentry tərəfindən təklif edilmişdi [32]. Homomorf şifrləmə sistemləri fərdi məlumatların qorunması, bulud hesablamaları, elektron səsvermə, maliyyə məlumatlarının emalı, tövsiyə sistemlərində tətbiq edilə bilər.

DP metodları fərdi konfidensiallığı qorumaq üçün verilənlərin faktiki qiymətlərinə normal paylanmadan olan təsadüfi qiymətlər əlavə edir (riyazi gözləməsi sıfır olan). Nəzərə almaq vacibdir ki, verilənlərin modifikasiyası nəticəsində verilənlər bazasının funksionallığı deqradasiya edir. Bu yanaşmanın problemlərindən biri konfidensiallıq ilə nəticələrin dəqiqliyi arasında olan balansdır. SMC yanaşmasının DP üzərində üstünlüyü nəticələrin təxmini deyil, dəqiq olmasıdır, lakin o, xeyli hesablamalar və hər bir təhlükəsiz kommunikasiya addımında əlavə kommunikasiya yükü tələb edir.

VI. DEEP LEARNING METODLARI

Çoxsaylı dərin neyron şəbəkə arxitekturalarından ən çox tətbiq ediləni konvolyusiya neyron şəbəkələridir (Convolutional Neural Network, CNN). CNN irəli yayılan çoxlaylı neyron şəbəkəsidir, şəkillərin tanınması və klassifikasiyası üçün geniş istifadə edilir [6]. Onun əsas tikinti blokları növbələşən konvolyusiya layları, pulinq layları, aktivləşdirmə funksiyası və tam əlaqəli laydır. Konvolyusiya layında lokal əlaqələndirmə və çəki paylaşımı yolu ilə siqnalın bəzi əsas xarakteristikaları – aşağı səviyyə əlamətləri çıxarılır. Konvolyusiyadan alınmış əlamət vektorunun ölçüsü böyük olur, onu azaltmaq üçün pulinq layları istifadə edilir. Bu laylardan sonra istifadə edilən tam əlaqəli lay lokal əlamətləri birləşdirir və konkret məsələlər üçün qlobal əlamətlər formalaşdırır. Bu laydakı hər bir neyron əvvəlki layın bütün neyronları ilə əlaqəlidir, lakin həmin layın neyronları arasında əlaqələr yoxdur.

CNN səhvin geriyə yayılması alqoritmi ilə öyrədilir, bu alqoritmdə stoxastik qradiyent enişi (Stochastik Gradient Descent, SGD) optimallaşdırma alqoritminin müxtəlif modifikasiyaları istifadə edilir.

Adətən, aktivləşdirmə funksiyası kimi ReLU (Rectified Linear Unit) istifadə edilir: $f(x) = max(0; x)$.

Dropout layı təlim zamanı overfittinq problemini həll etməyə xidmət edir [33]. Təlim zamanı dərin neyron şəbəkəsi təlim verilənlərində yüksək nəticə göstərir, lakin təlimdə rast gəlmədiyi test verilənlərində nəticələri xeyli pisləşir. Dropout metodunun mahiyyəti təlim prosesində layın seçilməsi və oradan müəyyən sayda (məsələn, 20 %) neyronun təsadüfi çıxarılmasıdır, onlar sonrakı hesablamalarda iştirak etmir. Belə üsul öyrənmənin effektivliyini və nəticənin keyfiyyətini yaxşılaşdırır.

Avtoenkoder çıxışda girişin özünü verməyi öyrənən irəli yayılan neyron şəbəkəsidir [6]. Giriş-gizli lay hissəsi enkoderə, gizli lay-çıxış hissəsi dekoderə uyğun gəlir. Dərin avtoenkoder avtoenkoderin enkoder və dekoderini bir neçə gizli laylarla genişləndirməklə qurulur. Küysüzləşdirən (ing. denoising) avtoenkoder – avtoenkoderə stoxastik küy əlavə etməklə daha robast əlamətlər öyrənməyə məcbur edilir.

VII. FƏRDİ MƏLUMATLARIN GİZLİLİYİ ÜÇÜN DEEP LEARNING YANAŞMALARI

[8]-də fərdi məlumatların konfidensiallığının təmin edilməsi üçün mövcud yanaşmaları homomorf şifrləmə, diferensial gizlilik və təhlükəsiz çoxtərəfli hesablama əsaslı olmaqla üç qrupa bölür.

[34]-də avtoenkoderlər əsasında dPA (ddep Private auto-encoder) metodu təklif edilir. Əsas ideya ənənəvi dərin enkoderin nəticəsini deyil, məqsəd funksiyasını perturbasiya etməklə ε-diferensial gizliliyi təmin etməkdir.

[35]-də ikisəviyyəli arxitektura təklif edir, arxitekturada modifikasiya edilmiş seyrək küysüzləşdirən avtoenkoder (verilənlərin çevrilməsi üçün) və dərin CNN şəbəkəsi (çevrilmiş verilənlərin klassifikasiyası üçün) istifadə edilir.

R. Shokri və V. Shmatikov dərin neyron şəbəkəsinin paylanmış təlimi üçün sistem yaradılır və qiymətləndirilir [36]. Bu sistemdə müasir təlim təlim metodlarında istifadə edilən stoxastik qradiyent enişi alqoritmlərinin paralelləşdirilməsi və asinxron yerinə yetirilməsi faktından istifadə edilir. Sistem iştirakçılara neyron şəbəkə modelini öz gizli verilənləri ilə öyrətməyə və təlim zamanı modelin əsas parametrlərinin kiçik altçoxluğunu selektiv paylaşmağa imkan verir. Bu dəqiqlik/konfidensiallıq fəzasında cəlbedici nöqtədir: iştirakçılar gizli verilənlərin konfidensiallığını təmin edirlər və digər iştirakçıların modellərindən faydalanaraq öz modellərinin öyrənmə dəqiqliyini artırırlar.

[37]-də piksel-əsaslı şəkil şifrləməsi metodu dərin neyron şəbəkələrində konfidensiallığı qorumaq üçün təklif edilir və ResNet-18 şəbəkəsi ilə şəkillərin klassifikasiyası üçün tətbiq edilir. Digər oxşar metodlardan fərqli olaraq, təklif edilmiş metod şifrlənmiş fəzada verilənlərin artırılmasına imkan verir.

H. Chabanne və həmmüəllifləri CryptoNets yanaşmasının (Microsoft tədqiqatçıları təklif etmişdir) dərin neyron şəbəkələrində (2-dən çox lay olduqda) qeyri-effektiv işləməsi problemini həll edirlər [38]. Bunun üçün CryptoNets-in orijinal ideyaları [39] ilə paket normallaşdırması ideyası birləşdirilir. Verilənlərin konfidensiallığını və emalın effektivliyini təmin etmək üçün CryptoNets-in əsas ideyası neyron şəbəkə ilə tam homomorf şifrləməni birləşdirməkdir. Təklif edilən yanaşma 6 qeyri-xətti layı olan neyron şəbəkəyə tətbiq edildikdə MNİST bazasında CryptoNets-i əhəmiyyətli dərəcədə yaxşılaşdırır.

[40]-da ReLU aktivləşmə funksiyası ilə CNN istifadə edilir və pulinq layı addımı artırılmış konvolyusiya layı ilə əvəz

edilir. Bu CNN-in strukturunu sadələşdirir və onun effektivliyini artırır. Konfidensiallığı təmin etmək üçün FHE istifadə edilir, lakin FHE hesablamalar baxımından effektiv deyil, buna görə CNN-i modifikasiya etmək lazım gəlir.

Neyron şəbəkələri tibbi və maliyyə məlumatları kimi konfidensial verilənlərə tez-tez tətbiq edilir.

Shokri və Shmatikov metodunda lokal verilənlər haqqında informasiyanın dürüst, amma maraqlı serverə sızdırıla biləcəyini göstərirlər və serverə informasiya sızıntısını aradan qaldırmaq üçün asinxron stoxastik qradiyent enişində addıtiv homomorf şifrləmə istifadə edirlər [41].

[42]-də çoxlaylı, məqsədi qeyri-qabarıq və parametrlərinin sayı olduqca çox olan modellərə baxılır. Neyron şəbəkələrinin diferensial gizli təliminə təklif edilən yanaşmanın əsas elementləri diferensial gizli SGD alqoritmi, momentlərin uçotu və hiperparametrlərin köklənməsi komponentləridir. Təklif edilən yanaşma TensorFlow əsasında reallaşdırılıb.

Momentlərin uçotu konfidensiallıq itkisini izləmək üçün istifadə edilir. Konfidensiallıq itkisi alqoritmə əlavə edilən təsadüfi küydən asılı olan təsadüfi kəmiyyətdir. Hər bir addımda konfidensiallıq itkisi momentlərinin loqarifmləri additiv toplanır.

Hiperparametrlərin köklənməsi konfidensiallıq, dəqiqlik və məhsuldarlığı balanslaşdırmaq, xüsusən də modelin konfidensiallıq ilə relevant xarakteristikaları üçün nəzərdə tutulub. Müəlliflər eksperimentlər vasitəsilə müşahidə edirlər ki, modelin dəqiqliyi neyron şəbəkəsinin strukturuna nisbətən, paketin ölçüsü və küy səviyyəsi kimi təlim parametrlərinə daha həssasdır.

[43]-də dərin neyron şəbəkələrinin şifrlənmiş verilənlər üzərində işləməsi üçün CryptoDL yanaşması təklif edilir, onun əsas komponentləri CNN və FHE-dir. CNN-də geniş istifadə edilən aktivləşdirmə funksiyalarının (ReLU, sigmoid, tanh) aşağı dərəcəli çoxhədlilərlə approksimasiyası metodu işlənir, bu effektiv homomorf şifrləmə sxümləri üçün olduqca vacibdir. Sonra laylarının sayı və strukturu dəyişən müxtəlif CNN-lərdə yeni aktivləşmə funksiyaları ilə MNIST aparılan eksperimentlər yanaşmanın düzgünlüyünü təsdiqləyir (99,52%, ən yaxşı nəticə 99,77 %-dir).

PrivyNet sensitiv təlim verilənləri probleminə diqqət yetirir [44]. Lokal dayaz neyron şəbəkələri qurulur və aralıq təsvirləri buludla paylaşaraq yekun model öyrədilir. Diferensial gizliliyi aralıq təsvirlərə küy əlavə etməklə təmin etməyə çalışırlar. Lokal dayaz neyron şəbəkələri açıq məlumatlardır və PrivyNet metodu MIA və GAN hücumlarına həssasdır.

NƏTİCƏ

İnformasiya və kommunikasiya texnologiyalarının sürətli inkişafı nəticəsində böyük həcmdə generasiya edilən fərdi məlumatların təhlükəsizliyinə böyük təhdidlər yaranır. Bu təhdidlərin bir qismi fərdi məlumatların intellektual analizi ilə əlaqədar meydana çıxır. Müasir intellektual analiz metodları müxtəlif mənbələrdə toplanmış böyük həcmli fərdi verilənləri əlaqələndirməklə və analiz etməklə fərdlər və sosial qruplar barəsində onların özünə də məlum olmayan yeni sensitiv biliklər aşkarlamağa imkan verir. Bu işdə intellektual analiz zamanı fərdi məlumatların konfidensiallığını təmin etmək üçün Deep Learning metodları ilə diferensial gizlilik, homomorf şifrləmə və təhlükəsiz çoxtərəfli hesablamaların birgə tətbiqi sahəsində mövcud işlərin icmalı verilmişdir. Müəllifin gəldiyi nəticəyə görə, diferensial gizlilik metodları ilə Deep Learning metodlarının kombinasiyası dəqiqlik və məhsuldarlıq balansının təmin olunması baxımından daha perspektivlidir.

MİNNƏTDARLIQ

Bu iş Azərbaycan Respublikası Dövlət Neft Şirkətinin (SOCAR) Elm Fondunun mailyyə yardımı ilə yerinə yetirilmişdir – **Müqavilə № 23LR – AMEA**.

İSTİNADLAR

[1] Y. İmamverdiyev “Big data və fərdi məlumatların təhlükəsizliyi,” Big data: imkanları, multidissiplinar problemləri və perspektivləri I respublika elmi-praktiki konfransı, s. 109-113, 2016.

[2] C.C. Aggarwal, and P.S. Yu. Privacy-preserving data mining: models and algorithms. New York: Springer, 2008.

[3] P. R. M. Rao, S. M. Krishna, & A. S. Kumar, “Privacy preservation techniques in big data analytics: a survey,” Journal of Big Data, 5(1):33, 2018. https://doi.org/10.1186/s40537-018-0141-8

[4] K. Kenthapadi, I. Mironov, & A. Thakurta “Privacy-preserving Data mining in industry,” Companion Proceedings of The 2019 World Wide Web Conference, pp. 1308-1310, 2019.

[5] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, 2015.

[6] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT Press, 2016.

[7] P. Xie, B. Wu, & G. Sun “BAYHENN: Combining bayesian deep learning and homomorphic encryption for secure DNN inference,” arXiv preprint arXiv:1906.00639. 2019.

[8] H. C. Tanuwidjaja, R. Choi, & K. Kim “A survey on deep learning techniques for privacy-preserving,” International Conference on Machine Learning for Cyber Security, pp. 29-46, 2019.

[9] D. Zhang, X. Chen, D. Wang, & , J. Shi “A survey on collaborative deep learning and privacy-preserving,” IEEE Third International Conference on Data Science in Cyberspace (DSC), pp. 652-658, 2018.

[10] A. Gkoulalas-Divanis, G. Loukides, & J. Sun “Publishing data from electronic health records while preserving privacy: A survey of algorithms,” Journal of biomedical informatics, 50, pp. 4-19, 2014.

[11] R. Shokri, M. Stronati, C. Song, V. Shmatikov “Membership inference attacks against machine learning models,” IEEE Symposium on Security and Privacy, pp.3-18, 2017.

[12] C. Dwork, & A. Roth “The algorithmic foundations of differential privacy,” Foundations and Trends in Theoretical Computer Science, 9(3–4), pp. 211-407,2014.

[13] N. Hamza, & H. A. Hefny “Attacks on anonymization-based privacy-preserving: a survey for data mining and data publishing,” Journal of Information Security, 4(02), pp. 101-112, 2013.

[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville & Y. Bengio, “Generative adversarial nets,” Advances in neural information processing systems, 2014, pp. 2672-2680.

[15] M. Fredrikson, S. Jha, & T.Ristenpart “Model inversion attacks that exploit confidence information and basic countermeasures,” Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, pp. 1322-1333, 2015.

[16] F. F. Zhao, L. F. Dong, K. Wang, Y. Li, “Study on privacy protection algorithm based on k-anonymity,” Physics Procedia, Vol.33, pp. 483-490, 2012.

[17] X. Qi, & M. Zong “An overview of privacy preserving data mining,” Procedia Environmental Sciences, 12, pp. 1341-1347, 2012.

[18] S. Upadhyay, C. Sharma, P. Sharma, P. Bharadwaj, & K. R. Seeja “Privacy preserving data mining with 3-D rotation transformation,” Journal of King Saud University-Computer and Information Sciences, 30(4), pp. 524-530, 2018.

[19] Z. Huang, W. Du, & B. Chen “Deriving private information from randomized data,” Proceedings of the 2005 ACM SIGMOD international conference on Management of data, pp. 37–48, 2005.

[20] S. E. Fienberg, & J. McIntyre “Data swapping: Variations on a theme by dalenius and reiss,” International Workshop on Privacy in Statistical Databases, pp. 14-29, 2004.

[21] L. Sweeney, “k-anonymity: A model for protecting privacy,” International Journal of Uncertainty Fuzziness and Knowledge Based Systems, vol. 10, no. 5, pp. 557-570, 2002.

[22] J. Dowd, S. Xu, & W. Zhang “Privacy-preserving decision tree mining based on random substitutions,” International Conference on Emerging Trends in Information and Communication Security, pp. 145-159, 2006.

[23] C.C. Aggarwal, Data mining: the textbook. Springer, 2015.

[24] C. Dwork, F. McSherry, K. Nissim, A. Smith “Calibrating noise to sensitivity in private data analysis,” Theory of Cryptography Conference, pp. 265–284, 2006.

[25] K. Nissim, T. Steinke, A. Wood, M. Altman, et al “Differential privacy: A primer for a non-technical audience,” Vanderbilt Journal of Entertainment & Technology Law, 21 (1), pp. 209-276, 2018.

[26] C. Dwork, & A. Roth “The algorithmic foundations of differential privacy,” Foundations and Trends in Theoretical Computer Science, 9(3–4), pp. 211-407, 2014.

[27] N. Papernot, M. Abadi, U. Erlingsson, I. Goodfellow, & K. Talwar “Semi-supervised knowledge transfer for deep learning from private training data,” arXiv preprint arXiv:1610.05755, 2016.

[28] A.C.-C. Yao “How to generate and exchange secrets,” Foundations of Computer Science 27th Annual Symposium, pp. 162–167, 1986.

[29] O. Goldreich, S. Micali, A. Wigderson “How to play any mental game,” Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing, pp. 218–229, 1987.

[30] Y. Lindell, B. Pinkas “Privacy preserving data mining,” In: Bellare M. (eds) Advances in Cryptology – CRYPTO 2000. Lecture Notes in Computer Science, vol 1880. 2000.

[31] R. L. Rivest, L. Adleman, & M. L. Dertouzos “On data banks and privacy homomorphisms,” Foundations of secure computation, 4(11), pp. 169-180, 1978.

[32] C. Gentry “Fully homomorphic encryption using ideal lattices,” Annual ACM on Symposium on Theory of Computing, pp. 169–178, 2009.

[33] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov “Dropout: A simple way to prevent neural networks from overfitting,” The Journal of Machine Learning Research, 15(1), pp. 1929–1958, 2014.

[34] N. Phan, Y. Wang, X. Wu, & D. Dou “Differential privacy preservation for deep auto-encoders: an application of human behavior prediction,” Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), pp. 1309-1316, 2016.

[35] R. M. Alguliyev, R. M. Aliguliyev, & F. J. Abdullayeva “Privacy-preserving deep learning algorithm for big personal data analysis,” Journal of Industrial Information Integration, 15, pp. 1-14, 2019.

[36] R. Shokri, & V. Shmatikov “Privacy-preserving deep learning,” Proceedings of the 22nd ACM SIGSAC conference on computer and communications security, pp. 1310-1321, 2015.

[37] W. Sirichotedumrong, T. Maekawa, Y. Kinoshita, & H. Kiya “Privacy-preserving deep neural networks with pixel-based image encryption considering data augmentation in the encrypted domain,” arXiv preprint arXiv:1905.01827, 2019.

[38] H. Chabanne, A. de Wargny, J. Milgram, C. Morel, & E. Prouff “Privacy-preserving classification on deep neural network,” IACR Cryptology ePrint Archive, 2017. https://ia.cr/2017/035

[39] R. Gilad-Bachrach et al., “Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy,” in Proc. 33nd Int. Conf. Mach. Learn. (ICML), vol. 48, pp. 201–210, 2016.

[40] W. Liu, F. Pan, X. A. Wang, Y. Cao, & D. Tang “Privacy-preserving all convolutional net based on homomorphic encryption,” International Conference on Network-Based Information Systems, pp. 752-762, 2018.

[41] L. T. Phong, Y. Aono, T. Hayashi, L. Wang, & S. Moriai “Privacy-preserving deep learning via additively homomorphic encryption,” IEEE Transactions on Information Forensics and Security, 13(5), pp. 1333-1345, 2018.

[42] M. Abadi , A. Chu , I. Goodfellow , H. B. McMahan , I. Mironov , K. Talwar , L.Zhang “Deep learning with differential privacy,” Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp. 24-28, 2016.

[43] E. Hesamifard, H. Takabi, M. Ghasemi “CryptoDL: deep neural networks over encrypted data,” arXiv preprint, arXiv:1711.05189, 2017.

[44] M. Li, L. Lai, N. Suda, V. Chandra, & D. Z. Pan “PrivyNet: A flexible framework for privacy-preserving deep neural network training,” arXiv preprint arXiv:1709.06161, 2017.

DEEP LEARNING METHODS FOR PRIVACY PRESERVING DATA MINING

Yadigar Imamverdiyev

Institute of Information Technology of ANAS, Baku, Azerbaijan

yadigar@iit.science.az

Abstract – The paper provides a survey of privacy preserving data mining methods, including deep learning. In recent decades, personal data has become one of the main goals of cybercriminals, and despite measures to ensure confidentiality of data, data mining methods allow to extract sensitive information about individuals. To counter this privacy threat, numerous modifications to traditional data mining methods have been proposed that work with sensitive personal data. This study analyzes the use of deep learning methods for privacy preserving, as well as the problems that arise when combining deep learning methods with differential privacy, secure multiparty computing and homomorphic encryption methods.

Keywords *– personal data, privacy, privacy threats, PPDM; deep learning*.