Federated Generative Models

مدل های مولد فدرال

فرزانه الشریف[۱]،شهرام محمدی[۲]،لیلا حسین زاده[۳]

[۱]دانشکده مهندسی کامپیوتر،واحد اصفهان،دانشگاه ملی مهارت دختران،اصفهان،ایران

[۲]دانشکده مهندسی کامپیوتر،واحد اصفهان،دانشگاه ملی مهارت دختران،اصفهان،ایران

[۳]دانشکده مهندسی کامپیوتر،واحد اصفهان،دانشگاه ملی مهارت دختران،اصفهان،ایران

چکیده

با پیشرفت فناوری‌های حفاظت از حریم خصوصی، یادگیری فدرال (FL[۱]) به عنوان یک راهکار اصلی و انتخابی برای بسیاری از سناریوهای کاربردی محاسبات محرمانه تبدیل شده است.

یادگیری فدرال (FL) یک پارادایم توزیع‌شده امن یادگیری ماشین است که به مسئله سیلوهای داده در ساخت یک مدل مشترک می‌پردازد. حالت آموزشی توزیع‌شده منحصر به فرد آن و مزایای مکانیسم جمع‌آوری امن، آن را برای کاربردهای عملی مختلف با الزامات شدید حفظ حریم خصوصی بسیار مناسب می‌کند.

فدرال می‌تواند با ارتباط بین کلاینت و سرور، یک مدل جهانی با عملکرد برتر به دست آورد.

مدل‌های مولد طوری طراحی شده‌اند که توزیع یک مجموعه داده را یاد بگیرند و نمونه‌های داده جدیدی تولید کنند که مشابه داده‌های اصلی باشند.استفاده همزمان از یادگیری فدرال و مدل‌های مولد می‌تواند مستعد حملات باشد و طراحی معماری بهینه همچنین چالش‌برانگیز است.

کلمات کلیدی : یادگیری فدرال، مدل های مولد ، حریم خصوصی ، داده

Federated learning[۱](FL)

مقدمه

مدل‌های مولد عمیق (یک دسته از مدل‌های یادگیری ماشین) طوری طراحی شده‌اند که توزیع احتمال زیربنایی یک مجموعه داده را یاد بگیرند و نمونه‌های داده جدیدی تولید کنند که مشابه داده‌های اصلی باشند. بسیاری از مدل‌های مولد مانند GAN[۱]ها، VAE[۲]ها و مدل‌های انتشار در طول این سال‌ها استفاده شده‌اند.

با افزایش محبوبیت یادگیری ماشین مدل‌هایی که روی بسیاری از دستگاه‌ها بدون اشتراک‌گذاری فایل‌های محلی کار می‌کنند، یادگیری فدرال محبوب شده است.

یادگیری فدرال یک روش جدید است که به جای جمع‌آوری تمام داده‌ها در یک مکان مرکزی، مدل را مستقیماً روی دستگاه‌های کاربران آموزش می‌دهد. این روش به حفظ حریم خصوصی داده‌ها کمک می‌کند و برای کاربردهایی که حجم داده‌ها بسیار زیاد است و یا انتقال داده‌ها هزینه بالایی دارد، مناسب است.

در سال‌های اخیر، بسیاری از کارهای قبلی تلاش کرده‌اند مدل‌های مولد را در تنظیمات یادگیری فدرال ادغام کنند تا داده‌های حساس را محافظت کنند و با جلوگیری از اشتراک‌گذاری داده‌های خام با سرورهای مرکزی، عملکرد مدل را افزایش دهند. FL[۳]و مدل‌های مولد می‌توانند به روش‌های مختلف با هم کار کنند. ما می‌توانیم تمام تحقیقات مرتبط را به سه گروه تقسیم کنیم: اول، مدل‌های مولد می‌توانند به صورت فدرال کار کنند. این بدان معناست که ما می‌توانیم فرآیند مولد را در سیستم‌های توزیع‌شده تسهیل کنیم بدون اینکه داده‌های خام محلی را با سرورهای مرکزی به اشتراک بگذاریم. دوم، مدل‌های مولد می‌توانند به مدل‌های FL حمله کنند یا از آن‌ها محافظت کنند. سوم، می‌توان از مدل‌های مولد برای بهبود عملکرد مدل‌های FL استفاده کرد.

یادگیری فدرال یک الگوی جدید در ML[۴] است که هدف آن تسهیل آموزش مدل‌های باکیفیت از طریق هماهنگی چندین مشتری یا دستگاه، همگی در حالی که حریم خصوصی داده‌های محلی مربوط به آن‌ها حفظ می‌شود. چارچوب بنیادین FL در ابتدا توسط تیم گوگل پیشنهاد شد و از آن زمان به بعد محبوبیت فزاینده‌ای در بین محققان پیدا کرده است. دلیل اصلی این امر توانایی ذاتی آن در ارائه حفاظت بیشتر از حریم خصوصی نسبت به رویکردهای سنتی ML است.

---

[۱] Generative Adversarial Network (GAN)

[۲] Variational Autoencoder (VAE)

[۳] Federated learning (FL)

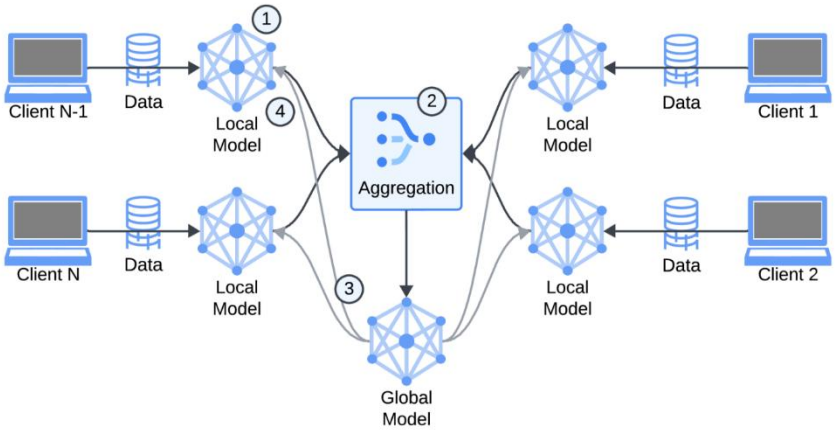
[۴] Machine learning (ML)

یادگیری فدرال

یادگیری فدرال یک فناوری قدرتمند است که می‌تواند در حوزه‌های مختلفی مانند پزشکی، مالی، شهر هوشمند و امنیت سایبری استفاده شود. با این حال، این فناوری با چالش‌های مختلفی نیز روبرو است. در این روش، مدل‌های یادگیری ماشین را می‌توان روی دستگاه‌های مختلف (مثل گوشی‌های موبایل) آموزش داد، بدون اینکه اطلاعات شخصی کاربران به یک سرور مرکزی ارسال شود. برای این کار، نیاز است که مدل‌های آموزش‌دیده روی دستگاه‌های مختلف، با هم ترکیب شوند تا یک مدل کلی ایجاد شود.

مدل های مولد فدرال

مدل‌های مولد فدرال **(Federated Generative Models)** یعنی نوعی از مدل‌های هوش مصنوعی که توانایی تولید داده‌های مصنوعی (مانند تصاویر، متن، ویدیو، یا هر نوع داده دیگر) را دارند، اما این کار را به شکلی انجام می‌دهند که داده‌های واقعی کاربران هرگز از دستگاه‌های آن‌ها خارج نشود.



شکل ۱. نمای کلی یادگیری فدرال

شکل ۱ نمایش بصری از فرایند یادگیری فدرال ارائه می‌دهد. در این شکل، چندین کلاینت (مثلاً دستگاه‌های مختلف) وجود دارند که هر کدام داده‌های محلی خود را دارند. هر کلاینت یک مدل محلی را بر روی داده‌های خود آموزش می‌دهد. سپس، این مدل‌های محلی به یک سرور مرکزی ارسال می‌شوند و در آنجا با هم ترکیب می‌شوند تا یک مدل جهانی ایجاد شود. این مدل جهانی به روز شده مجدداً به کلاینت‌ها بازگردانده می‌شود تا آن‌ها بتوانند مدل‌های محلی خود را بر اساس آن به روز کنند. این فرآیند به صورت تکراری انجام می‌شود تا زمانی که مدل به عملکرد مطلوب برسد.

اجزای کلیدی شکل:

کلاینت‌ها (Client) : هر یک از دستگاه‌ها یا سیستم‌هایی که داده‌های محلی خود را دارند و مدل‌های محلی را آموزش می‌دهند.

داده‌های محلی (Local Data) : داده‌هایی که در هر کلاینت ذخیره شده‌اند و برای آموزش مدل‌های محلی استفاده می‌شوند.

مدل‌های محلی (Local Model) : مدل‌های یادگیری ماشینی که بر روی داده‌های محلی هر کلاینت آموزش داده می‌شوند.

سرور مرکزی (Central Server) : سروری که مدل‌های محلی را جمع‌آوری کرده، آن‌ها را با هم ترکیب می‌کند و مدل جهانی را به روز شده به کلاینت‌ها باز می‌گرداند.

مدل جهانی (Global Model) : مدلی که از ترکیب مدل‌های محلی ایجاد می‌شود و نماینده دانش مشترک بین تمام داده‌ها است.

مزایا یادگیری فدرال

- حفظ حریم خصوصی: داده‌های حساس در دستگاه‌های محلی باقی می‌مانند و به سرور مرکزی منتقل نمی‌شوند.
- انعطاف‌پذیری: می‌توان از آن برای انواع مختلف داده‌ها و مدل‌های یادگیری ماشین استفاده کرد.
- مقیاس‌پذیری: می‌تواند برای آموزش مدل‌ها روی داده‌های بسیار بزرگ و توزیع‌شده استفاده شود.
- کاهش هزینه: از آنجایی که داده‌ها در دستگاه‌های محلی پردازش می‌شوند، نیازی به انتقال حجم زیادی از داده‌ها به سرور مرکزی نیست.

کاربرد یادگیری فدرال

با تلاش‌های مشترک بسیاری از محققان، FL[1] نقش مهمی در زمینه‌های مختلف صنایع ایفا کرده است. و کاربردهای اصلی فعلی FL در حوزه‌های بهداشت، مالی، صنعت و خدمات شهری است. در این بخش، ما کاربرد FL را در این سناریوهای کلاسیک به تفصیل شرح می‌دهیم.

در **حوزه پزشکی**، داده‌های بیمار داده‌های حساسی هستند و بسیاری از بیماران تمایل ندارند داده‌های خصوصی خود را برای خدمات آزمایش هوشمند به اشتراک بگذارند. در این شرایط، امنیت و کارایی فناوری FL پیشنهادهای عملی را برای کاربرد خدمات هوشمند در حوزه پزشکی ارائه می‌دهد. ترکیب FL و صنعت پزشکی می‌تواند بدون آسیب رساندن به اطلاعات خصوصی بیمار، مدل تشخیص بیماری با عملکرد خوبی ایجاد کند که فرآیند هوشمندسازی صنعت پزشکی را ارتقا می‌دهد.

موبایل : توسعه برنامه‌های کاربردی هوشمند بر روی دستگاه‌های تلفن همراه بدون نیاز به ارسال داده‌های شخصی کاربران به سرور.

اینترنت اشیا : آموزش مدل‌های یادگیری ماشین برای دستگاه‌های متصل به اینترنت بدون نیاز به انتقال داده‌های حسگرها به یک مکان مرکزی.

---

[1] Federated learning (FL)

شهر هوشمند : توسعه شهرها جدایی‌ناپذیر از تلاش‌های مشترک سازمان‌های دولتی، شرکت‌های خصوصی و افراد است و بهبود الزامات حفاظت از حریم خصوصی منجر می‌شود که مراکز داده نتوانند داده‌ها را به دلخواه در اختیار طرف‌های ثالث قرار دهند. این امر تبادل داده در سطوح مختلف را محدود می‌کند. ظهور یادگیری فدرال به طور موثر مشکل جزیره‌ای شدن داده‌ها را حل می‌کند و داده‌ها را در تمام سطوح شهر ادغام می‌کند تا خدمات شهری بهتر و خدمات حمل‌ونقل را برای شهروندان فراهم کند.

امور مالی و بیمه : کاربرد FL در سناریوهای مالی شامل کمک به بانک‌های مختلف در پیش‌بینی امتیاز اعتباری سپرده‌گذاران، پیش‌بینی توانایی بازپرداخت وام و کمک به موسسات مالی مختلف در درک بهتر توانایی سرمایه‌گذاری مشتریان و امتیازدهی اعتباری است.

چالش‌های یادگیری فدرال

در حالی که FL پتانسیل زیادی دارد، هنوز چالش‌های فنی و عملی زیادی وجود دارد که باید برطرف شود. برخی از مهم‌ترین چالش‌ها عبارتند از:

- **ناهمگنی آماری**: داده‌های تولید شده توسط دستگاه‌های مختلف اغلب دارای توزیع‌های متفاوتی هستند که می‌تواند بر عملکرد مدل‌های FL تأثیر بگذارد.
- **گلوگاه‌های ارتباطی**: در سیستم‌های بزرگ مقیاس FL، ارتباط بین دستگاه‌ها و سرور مرکزی می‌تواند یک گلوگاه باشد.
- **جمع‌آوری ایمن**: حفظ حریم خصوصی داده‌ها در FL یک چالش اساسی است.

انواع یادگیری فدرال

انواع مختلف یادگیری فدرال (FL) بر اساس تفاوت‌ها در اندازه ویژگی و نوع داده بین کلاینت‌ها و سرور مرکزی طبقه‌بندی می‌شوند.

متمرکز در مقابل غیرمتمرکز: این دسته به نحوه طراحی آموزش و جمع‌آوری اشاره دارد. متمرکزترین رویکرد رایج، از یک سرور مرکزی برای مدیریت مراحل مختلف آموزش و جمع‌آوری مدل در سراسر همه منابع داده محلی استفاده می‌کند. از سوی دیگر، یادگیری فدرال غیرمتمرکز (همچنین به عنوان همتا به همتا شناخته می‌شود) شامل هماهنگی فردی کلاینت‌ها با یکدیگر بدون وجود یک سرور مرکزی است. در این حالت، پارامترهای مدل از یک کلاینت به کلاینت دیگر در یک زنجیره منتقل می‌شوند.

افقی، عمودی و انتقال یادگیری: این نوع بر اساس نحوه تقسیم داده‌ها بین کلاینت‌های مختلف است. در یادگیری فدرال افقی (همچنین به عنوان همگن یا مبتنی بر نمونه شناخته می‌شود)، مجموعه داده‌های کلاینت‌های مختلف دارای ویژگی‌های یکسان اما همپوشانی کمی در فضای نمونه دارند. در مقابل، یادگیری فدرال عمودی (همچنین به عنوان ناهمگن یا مبتنی بر ویژگی شناخته می‌شود) برای مجموعه داده‌هایی استفاده می‌شود که حاوی مجموعه ویژگی‌های متفاوتی هستند .

انواع داده

سه نوع مختلف تقسیم‌بندی داده‌ها در فدرال شامل:

الف) **یادگیری فدرال افقی (HFL[1])**: در این نوع، داده‌های موجود در کلاینت‌های مختلف دارای ویژگی‌های یکسان اما نمونه‌های متفاوت هستند. به عبارت دیگر، فضای ویژگی و فضای برچسب در همه کلاینت‌ها یکسان است اما نمونه‌ها متفاوت هستند. این نوع یادگیری فدرال برای سناریوهایی مناسب است که در آن داده‌های هر کلاینت نماینده یک بخش از توزیع کلی داده‌ها است.

ب) **یادگیری فدرال عمودی (VFL[2])**: در این نوع، داده‌های موجود در کلاینت‌های مختلف دارای ویژگی‌های متفاوت اما نمونه‌های یکسانی هستند. به عبارت دیگر، فضای ویژگی در کلاینت‌های مختلف متفاوت است اما فضای برچسب یکسان است. این نوع یادگیری فدرال برای سناریوهایی مناسب است که در آن هر کلاینت دارای اطلاعات مکمل در مورد همان نمونه‌ها است.

ج) **یادگیری انتقال فدرال (FTL[3])**: این نوع یادگیری فدرال ترکیبی از یادگیری فدرال افقی و عمودی است. در این نوع، هدف انتقال دانش از یک حوزه (که داده‌های زیادی دارد) به حوزه دیگری (که داده‌های محدودی دارد) در تنظیمات یادگیری فدرال است.

روش های یادگیری فدرال:

۱. الگوریتم FedAvg[4] : این الگوریتم به یک روش استاندارد در یادگیری فدرال تبدیل شده است. فرآیند این الگوریتم در چهار مرحله انجام می‌شود:

۱. سرور مرکزی مدل جهانی را به سازمان‌ها ارسال می‌کند.

۲. هر سازمان مدل خود را با استفاده از داده‌های محلی به‌روزرسانی می‌کند.

۳. سازمان‌ها مدل‌های محلی به‌روزرسانی‌شده را به سرور ارسال می‌کنند.

۴. سرور وزن‌های مدل‌ها را میانگین‌گیری کرده و مدل جهانی را به‌روزرسانی می‌کند.

---

[1] Horizontal Federated Learning (HFL)
[2] Vertical Federated Learning (VFL)
[3] Federal Transfer Learning (FTL)
[4] Federated Averaging (FedAvg)

برای بهبود عملکرد FedAvg روی داده‌های ناهمگن، دو رویکرد اصلی وجود دارد:

بهبود آموزش محلی: الگوریتم‌هایی مانند FedProx و SCAFFOLD تلاش می‌کنند تا آموزش محلی را کنترل کنند. FedProx محدودیت‌هایی بر فاصله مدل محلی و مدل جهانی اعمال می‌کند، در حالی که SCAFFOLD از متغیرهای کنترلی برای تصحیح به‌روزرسانی‌ها استفاده می‌کند.

بهبود ادغام مدل‌ها: الگوریتم‌هایی مانند FedMA و FedNova روش‌هایی برای ادغام بهتر مدل‌های محلی ارائه داده‌اند.

۲. یادگیری متضاد:

یادگیری متضاد، نوعی یادگیری خودنظارتی است که بازنمایی داده‌ها را با استفاده از داده‌های بدون برچسب بهبود می‌دهد. این روش تلاش می‌کند فاصله میان بازنمایی‌های تصاویر مشابه را کاهش داده و بازنمایی تصاویر مختلف را از هم دور کند. یکی از روش‌های معروف در این حوزه، SimCLR است. MOON با الهام از این روش، یادگیری متضاد را در سطح مدل‌ها اجرا می‌کند.

۳. روش MOON:

MOON یک روش ساده اما مؤثر بر اساس FedAvg است. در هر دور از یادگیری:

- فاصله بین بازنمایی مدل محلی و مدل جهانی کاهش می‌یابد.
- فاصله بازنمایی مدل محلی با نسخه قبلی آن افزایش می‌یابد.

توزیع داده‌ها: IID و Non-IID

IID (مستقل و یکسان توزیع شده): در این حالت، فرض بر این است که داده‌های موجود در کلاینت‌های مختلف مستقل از یکدیگر هستند و از یک توزیع احتمال یکسان پیروی می‌کنند.

Non-IID: در این حالت نمونه‌های داده‌ای از یکدیگر مستقل نبوده و یا از توزیع‌های احتمالی متفاوتی پیروی می‌کنند.

رایج‌ترین مدل‌های مولد

شبکه‌های مولد متخاصم GAN :(GANs)ها نوعی از الگوریتم‌های یادگیری ماشینی هستند که شامل دو شبکه عصبی - مولد و تمایزگر - هستند که در یک بازی مین-مکس شرکت می‌کنند. هدف مولد تولید داده‌هایی است که به قدری واقع‌گرا باشد که تمایزگر نتواند آن‌ها را از داده‌های واقعی تشخیص دهد. از سوی دیگر، هدف تمایزگر طبقه‌بندی دقیق داده‌های واقعی از داده‌های جعلی تولید شده توسط مولد است. این بازی مین-مکس بهینه‌سازی به هر دو شبکه کمک می‌کند تا عملکرد خود را بهبود بخشند و منجر به تولید داده‌های مصنوعی واقع‌گرا شود.

اتوکدهای واریاسیونی (VAEs): اتوکدهای واریاسیونی نوعی از مدل‌های مولد هستند که از معماری اتوکدها برای ایجاد داده‌های با ابعاد بالا استفاده می‌کنند . آن‌ها شامل یک کدگذار هستند که داده‌های ورودی را به یک نمایش با ابعاد پایین‌تر فشرده می‌کند و یک رمزگشا که داده‌ها را از این نمایش بازسازی می‌کند.

مدل‌های انتشار: مدل‌های انتشار نوعی از مدل‌های مولد هستند که فرآیند تدریجی افزودن نویز و سپس حذف نویز را برای تولید داده‌ها شبیه‌سازی می‌کنند. این مدل با یک توزیع نویز خالص شروع می‌شود و به تدریج این نویز را به نمونه‌هایی از یک توزیع هدف تبدیل می‌کند. این رویکرد نمونه‌های بسیار دقیق و منسجمی تولید می‌کند و مدل‌های انتشار را به ویژه در تولید تصاویر و صدا با کیفیت بالا موثر می‌سازد.

حملات به فدرال

این حملات را می‌توان به دو دسته اصلی تقسیم کرد: حملات به حریم خصوصی و حملات به یکپارچگی.

حملات به حریم خصوصی: مانند بازسازی مدل، استنتاج و وارون‌سازی مدل.

نمونه هایی از حملات حریم خصوصی

حملات استنتاج عضویت: حملات استنتاج عضویت سعی می‌کنند کشف کنند که آیا یک نمونه داده به داده‌های آموزشی تعلق دارد یا خیر. برای مثال، یک حمله استنتاج عضویت می‌تواند فاش کند که آیا داده‌های یک بیمار خاص در آموزش مدل برای پیش‌بینی بیماری آلزایمر استفاده شده است یا خیر. اگر یک مهاجم بتواند تشخیص دهد که داده‌های یک بیمار خاص در مجموعه داده آموزشی مدل گنجانده شده است، می‌تواند استنباط کند که این بیمار در معرض خطر ابتلا به آلزایمر است یا قبلاً تشخیص داده شده است.

حملات استنتاج صفت: یک حمله استنتاج صفت می‌تواند به دنبال کشف ویژگی‌های خاص یا اطلاعات محرمانه در مورد افراد در یک مجموعه داده باشد. برای مثال، یک مهاجم می‌تواند با تجزیه و تحلیل الگوهای موجود در این داده‌ها، ویژگی‌های حساس مانند سن، موقعیت مکانی و عادات مرور وب را استنباط کند. هدف اصلی چنین حمله‌ای پیش‌بینی اطلاعات حساس در مورد افراد با استفاده از داده‌های موجود است .

حملات استنتاج ویژگی: استنتاج ویژگی زمانی رخ می‌دهد که بتوانیم زمانی را که یک ویژگی در مجموعه داده ظاهر و ناپدید می‌شود، شناسایی کنیم . برای مثال، یک حمله استنتاج ویژگی می‌تواند مشخص کند که آیا یک فرد خاص در عکس‌های استفاده شده برای آموزش یک طبقه‌بند تشخیص چهره حضور داشته است یا خیر. اگر یک مهاجم بتواند تشخیص دهد که یک فرد خاص در مجموعه داده آموزشی حضور داشته است، می‌تواند استنباط کند که این فرد به طور بالقوه در معرض شناسایی قرار دارد.

حملات بازسازی: در حملات بازسازی، یک مهاجم تلاش می‌کند تا داده‌های خصوصی اصلی کلاینت‌ها را از به‌روزرسانی‌های مدل به اشتراک گذاشته شده بازسازی کند. یک مهاجم از تکنیک‌های استنتاج و تحلیل آماری برای استخراج ویژگی‌های داده‌های اصلی از این به‌روزرسانی‌ها استفاده می‌کند.

حملات وارون‌سازی مدل: ایده اصلی پشت حملات وارون‌سازی مدل این است که از مزایای نحوه یادگیری و ذخیره اطلاعات توسط مدل‌های یادگیری ماشین استفاده کند. یک مهاجم می‌تواند با دستکاری مدل و مشاهده خروجی‌های آن، ویژگی‌های داده‌های استفاده شده برای آموزش مدل را استنباط کند . برای مثال، با ردیابی دقیق پرس‌وجوهای ورودی و مشاهده خروجی‌های مدل، یک مهاجم می‌تواند ویژگی‌های داده‌های استفاده شده برای آموزش مدل را استنباط کند. همچنین در برخی موارد، مهاجم ممکن است با ویژگی‌های استنباط شده داده‌ها، داده‌های اصلی را بازسازی کند.

حملات به یکپارچگی: مانند حملات پس‌در و مسموم کردن مدل.

نمونه هایی از حملات به یکپارچگی

مسوم سازی: در مکانیسم FL، حمله مسموم‌سازی تمایل دارد تا داده‌های آموزشی محلی مشتری یا مدل تولید شده در فرآیند آموزش FL را دستکاری، تخریب یا آلوده کند تا امنیت سیستم FL را به خطر بیندازد . انواع موجود مسموم‌سازی داده‌ها در FL معمولاً به مسموم‌سازی داده‌ها و مسموم‌سازی مدل تقسیم می‌شود که تهدید بزرگی برای امنیت FL است.

مسموم‌سازی داده‌ها: مهاجمان با دستکاری داده‌های آموزشی سعی می‌کنند مدل را فریب دهند.

مسموم‌سازی مدل: مهاجمان با دستکاری مدل‌های محلی، مدل جهانی را آلوده می‌کنند.

برای مقابله با این حملات، روش‌های مختلفی مانند حذف داده‌های پرت، استفاده از بلاکچین و تحلیل استثناهای فدرال پیشنهاد شده است.

تکنیک‌های حفاظت از حریم خصوصی در یادگیری فدرال

به عنوان یک فناوری یادگیری ماشینی امن با حفاظت از حریم خصوصی، درجه حفاظت از اطلاعات خصوصی توسط FL عمدتاً توسط فناوری‌های رمزنگاری تضمین می‌شود. در حال حاضر، تکنیک‌های رمزنگاری سنتی پذیرفته شده توسط اکثر محققان شامل محاسبات چندجانبه امن، حریم خصوصی دیفرانسیلی و رمزنگاری هومومورفیک می‌شود. در ادامه، این تکنیک‌های رمزنگاری مورد استفاده در FL به طور خلاصه معرفی می‌شوند.

محاسبات چندجانبه امن (SMC[1])

محاسبات چندجانبه امن برای محافظت از داده‌های ورودی هر طرف شرکت‌کننده در همکاری استفاده می‌شد. در این فرآیند، داده‌های حساس توسط رمزنگاری بین طرف‌ها محافظت می‌شود. سیستم محاسبات چندجانبه امن است که به کاربران اجازه می‌دهد داده‌ها را بدون دیدن آن پردازش کنند.

حریم خصوصی دیفرانسیلی (DP[۲])

پس از رمزگذاری داده‌های کاربر و آپلود آن به سرور مرکزی، سرور مخرب می‌تواند از داده‌های رمزگذاری شده برای استنباط ویژگی‌های گروه کاربر استفاده کند، اما نمی‌تواند اطلاعات دقیق یک فرد را تجزیه و تحلیل کند این روش حمله به حمله دیفرانسیلی معروف است.

با هدف مقابله با این تهدید، فناوری حریم خصوصی دیفرانسیلی (DP) می‌تواند از نویز تصادفی برای غرق کردن داده‌های اصلی استفاده کند و از این طریق مانع از بازگرداندن داده‌های اصلی از پایگاه داده شود. DP راهی برای محافظت از مشکل افزودن نویز به نتایج پرس و جو برای حل مسئله حفاظت از حریم خصوصی یک پرس و جو واحد است. DP دارای نظریه داده‌های دقیق است و ماهیت آن حفاظت از نتایج محاسباتی به جای فرآیند محاسباتی است. مزیت DP این است که دارای یک مدل حفاظت از حریم خصوصی است که کاملاً مستقل از دانش پس‌زمینه است و از نظر تئوری در برابر هرگونه حمله مقاوم است.

رمزنگاری همریخت (HE[۳])

رمزنگاری همریخت یک الگوریتم رمزنگاری است که خاصیت عملیات همریختی روی متن رمزنگاری شده را برآورده می‌کند. به عبارت دیگر، به کاربران اجازه می‌دهد مستقیماً عملیات جبری خاصی را روی متن رمزنگاری شده انجام دهند و نتیجه محاسبه روی متن رمزنگاری شده همان نتیجه‌ای است که پس از انجام همان عملیات روی متن ساده رمزنگاری شده به دست می‌آید.

طراحی یک الگوریتم HE کارآمد شامل برخی دانش رمزنگاری بر اساس نظریه پیچیدگی محاسبات ریاضی می‌شود. اگر یک الگوریتم HE از هر نوع محاسبه روی متن رمزنگاری شده پشتیبانی کند، به آن رمزنگاری کاملاً همریخت (FHE[۴]) می‌گویند. اگر از محاسبات جزئی روی متن رمزنگاری شده پشتیبانی کند، مانند جمع، ضرب یا تعداد محدودی از عملیات جمع، به آن رمزنگاری نیمه‌همریخت یا جزئی همریخت (PHE[۵]) می‌گویند.

---

۱ Secure Multi-party Computation (SMC)
۲ Differential Privacy (DP)
۳ Homomorphic Encryption (HEL)
۴ Fully Homomorphic Encryption (FHE)
۵ Partially Homomorphic Encryption (PHE)

نتیجه گیری

در این مقاله، یک بررسی جامع در مورد یادگیری فدرال (FL) و مدل های مولد انجام داده‌ایم. FL به عنوان یک الگوی یادگیری ماشین توزیع شده با حفاظت از حریم خصوصی پیشرفته ظهور کرده است و جمع‌آوری مدل نقش حیاتی در این زمینه ایفا می‌کند.

یادگیری فدرال و مدل‌های مولد دو حوزه تحقیقاتی فعال هستند که پتانسیل بالایی برای حل چالش‌های مختلف در زمینه هوش مصنوعی دارند. با این حال، هنوز چالش‌های زیادی در این زمینه وجود دارد که نیازمند تحقیقات بیشتر است.

یادگیری فدرال (FL) به عنوان یک پارادایم پیشرفته در یادگیری ماشین توزیع‌شده پرداخته است که با حفظ داده‌های کاربران به صورت محلی، امکان ایجاد مدل‌های جهانی بدون نیاز به انتقال اطلاعات حساس را فراهم می‌کند. فدرال علاوه بر بهبود امنیت داده‌ها، موجب کاهش هزینه‌های انتقال و افزایش انعطاف‌پذیری در مواجهه با داده‌های توزیع‌شده می‌شود و به‌ویژه در حوزه‌هایی نظیر سلامت، مالی و شهرهای هوشمند کاربرد دارد.

این فناوری قابلیت تولید داده‌های مصنوعی ایمن را نیز از طریق مدل‌های مولد فراهم می‌کند. با این حال، چالش‌های مهمی مانند ناهمگنی داده‌ها، محدودیت‌های ارتباطی بین دستگاه‌ها و سرورها و ریسک‌های مرتبط با امنیت همچنان به عنوان موانع موجود مطرح هستند. در عین حال، حملاتی نظیر بازسازی داده و مسموم کردن مدل‌ها، تهدیدی برای یکپارچگی و حریم خصوصی در یادگیری فدرال به شمار می‌آیند. فناوری‌هایی مانند رمزنگاری پیشرفته و محاسبات چندجانبه امن برای رفع این چالش‌ها پیشنهاد شده‌اند.

با وجود مزایای متعدد، تحقیقات نشان می‌دهند که برخی از فرضیات امنیتی FL ممکن است در برابر حملات مخرب کافی نباشد. بنابراین، توسعه الگوریتم‌های پیشرفته‌تر و تقویت حفاظت از حریم خصوصی برای گسترش ایمن این فناوری ضروری است.

منابع

[1] Pian Qi, Diletta Chiaro, Antonella Guzzo, Michele lanni, Giancarlo Fortino, Francesco Piccialli, Model aggregation techniques in federated learning: A comprehensive survey, Future Generation Computer Systems, Volume 150, 2024 Pages 272-2931-117-7397.

[2] Ashkan Vedadi Gargary, Emiliano De Cristofaro , 24 Pages, 3 Figures, 5 Tables , Machine Learning (cs.LG); Computation and Language (cs.CL); Cryptography and Security (cs.CR) , arXiv:2405.16682, Submitted on 26 May 2024.

[3] Jie Wen, Zhixia Zhang, Yang Lan, Zhihua Cui, Jianghui Cai, Wensheng Zhang ,Volume14,pages513-535,(2023), 11November2022.

[4] Tan, Yue and Long, Guodong and Ma, Jie and LIU, LU and Zhou, Tianyi and Jiang, Jing, Advances in Neural Information Processing Systems, Pages 19332- 19344 , Curran Associates, Inc,Federated Learning from Pre-Trained Models: A Contrastive Learning Approach, 2022.

[5] Qinbin Li, Bingsheng He, Dawn Song; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 10713-10722.

[6] Chen Zhang, Yu Xie, Hang Bai, Bin Yu, Weihong Li, Yuan Gao, A survey on federated 2021 , 106775, ISSN 0950-7051., Knowledge-Based Systems, Volume 216,learning