

طراحی و ارزیابی یک چارچوب یادگیری فدرال مبتنی بر بلاکچین برای اینترنت اشیا

۱. مقدمه

این پایان نامه بررسی می‌کند که چگونه فناوری بلاکچین می‌تواند از یادگیری فدرال (Federated Learning) یا FL در محیط‌های اینترنت اشیا که نسبت به حریم خصوصی حساس هستند، پشتیبانی کند.

به عنوان مطالعه موردی پیشنهادی — که انتخاب نهایی آن منوط به تأیید استاد راهنما است — این پژوهش موضوع نگهداری و تعمیرات پیش‌بینانه در اینترنت اشیا صنعتی (IIoT Predictive Maintenance) را بررسی می‌کند.

در چنین محیطی، اشتراک‌گذاری داده‌های خام حسگرها، مانند:

- لرزش؛
- دما؛
- نرخ جریان؛

بین کارخانه‌ها یا تأسیسات مختلف برای آموزش مدل‌های پیش‌بینی خرابی، ممکن است نگرانی‌هایی در زمینه حریم خصوصی داده‌ها و مالکیت فکری ایجاد کند.

معماری یادگیری فدرال افقی (Horizontal Federated Learning) یا HFL می‌تواند این مشکل را برطرف کند. در این معماری، دستگاه‌های لبه‌ای شبیه‌سازی‌شده، مدل‌های محلی خود را به صورت مستقل و با استفاده از داده‌های خود آموزش می‌دهند و فقط پارامترهای مدل را با یکدیگر مبادله می‌کنند. بنابراین، داده‌های خام دستگاه‌ها جابه‌جا یا به اشتراک گذاشته نمی‌شوند.

مطابق طرح مفهومی ارائه‌شده توسط استاد راهنما، کل محیط روی یک کامپیوتر شبیه‌سازی خواهد شد و شامل موارد زیر است:

- یک کنترل‌کننده یا Controller برای هماهنگ کردن فرایند؛
- سه دستگاه یا Device که هر کدام بخش جداگانه‌ای از داده‌های آموزشی را نگهداری می‌کنند؛
- استفاده از مدل رگرسیون لجستیک (Logistic Regression) برای طبقه‌بندی دودویی.

یک بلاکچین خصوصی و محلی اتریوم با استفاده از Geth از این فرایند پشتیبانی خواهد کرد. این کار از طریق قراردادهای هوشمند و با دو هدف اصلی تعیین‌شده در طرح استاد انجام می‌شود:

۱. شناسایی: ثبت Controller و دستگاه‌های شرکت‌کننده؛
۲. ثبت سوابق: ثبت اطلاعاتی مانند:
 - هویت شرکت‌کنندگان؛
 - اطلاعات دوره‌های آموزشی؛
 - زمان ثبت اطلاعات؛
 - هش به روزرسانی مدل‌ها.

این قابلیت‌ها امکان ردیابی سوابق تبادل اطلاعات و بررسی یکپارچگی و صحت به‌روزرسانی‌های مدل را فراهم می‌کنند.

نمودار صفحه اول نیز ارتباط میان کنترل‌کننده یادگیری فدرال، کلاینت‌ها، بخش‌های داده حسگر یا پمپ و بلاکچین خصوصی Geth را نشان می‌دهد.

۱,۱ اهداف

اهداف این پایان‌نامه عبارت‌اند از:

۱. طراحی و پیاده‌سازی یک محیط شبیه‌سازی شده یادگیری فدرال مبتنی بر بلاکچین یا BCFL روی یک کامپیوتر؛
۲. بررسی اینکه چگونه یک بلاکچین خصوصی اتریوم می‌تواند اعتماد، شناسایی و قابلیت ردیابی را در فرایند یادگیری فدرال فراهم کند؛
۳. ارزیابی عملکرد یادگیری ماشین و همچنین سریار ایجاد شده توسط بلاکچین در سناریوهای عملیاتی مختلف؛
۴. ارزیابی امکان استفاده از یادگیری فدرال مبتنی بر بلاکچین در کاربردهای اینترنت اشیا حساس به حریم خصوصی.

۲. رویکرد یادگیری فدرال

هر دور یا Round از فرایند یادگیری فدرال مطابق چرخه تکرار شونده تعریف شده در طرح استاد انجام می‌شود:

۱. کنترل‌کننده، مدل سراسری فعلی یا Global Model را برای دستگاه‌ها ارسال می‌کند.
۲. هر دستگاه مدل را به‌صورت محلی و با استفاده از داده‌های خصوصی خودش آموزش می‌دهد.
۳. دستگاه‌ها وزن‌ها یا پارامترهای به‌روز شده مدل را به کنترل‌کننده باز می‌گردانند.
۴. کنترل‌کننده به‌روزرسانی‌های دریافت شده را با یکدیگر جمع می‌کند و مدل سراسری جدید را ایجاد می‌کند.

برای تجمیع مدل‌ها از الگوریتم میانگین‌گیری فدرال یا Federated Averaging (FedAvg) استفاده می‌شود.

هر دستگاه مراحل آموزش محلی مبتنی بر گرادین کاهشی را با استفاده از مدل سراسری آغاز می‌کند. سپس کنترل‌کننده مدل‌های محلی ایجاد شده را بر اساس تعداد نمونه‌های موجود در هر دستگاه، به‌صورت وزن‌دار میانگین‌گیری می‌کند.

به زبان ساده، اگر یک دستگاه داده‌های بیشتری داشته باشد، تأثیر آن در ساخت مدل سراسری جدید بیشتر خواهد بود.

از آنجاکه داده‌های غیرهم‌توزیع و غیرمستقل یا Non-IID در محیط‌های اینترنت اشیا رایج هستند، ممکن است دستگاه‌ها داده‌هایی با ویژگی‌ها و توزیع‌های متفاوت داشته باشند.

برای مثال، در مطالعه موردی پیشنهادی، پمپ‌های مختلف ممکن است تحت شرایط بار کاری متفاوت فعالیت کنند.

یکی از توسعه‌های احتمالی پروژه، مقایسه الگوریتم پایه FedAvg با الگوریتم FedProx است.

FedProx با اضافه کردن یک عبارت تنظیم‌کننده، از فاصله گرفتن بیش از حد مدل محلی هر دستگاه از مدل سراسری جلوگیری می‌کند. در صورت تأیید استاد، این مقایسه می‌تواند اطلاعات بیشتری درباره نحوه همگرایی مدل در شرایطی ارائه دهد که داده‌های دستگاه‌ها ناهمگون هستند.

بااین‌حال، مدل محلی و سراسری در تمام مراحل همچنان رگرسین لجستیک باقی خواهد ماند. بنابراین، مقایسه فقط بر روش تجمیع تمرکز خواهد داشت و تمرکز اصلی پژوهش همچنان بر یکپارچه‌سازی بلاکچین و یادگیری فدرال باقی می‌ماند.

۳. طراحی قرارداد هوشمند

برای کاهش هزینه ذخیره‌سازی روی بلاکچین و کاهش تأخیر تراکنش‌ها، داده‌های خام حسگرها و وزن‌های کامل مدل خارج از بلاکچین یا Off-Chain نگهداری می‌شوند.

دستگاه‌ها فقط اطلاعات موردنیاز برای تأیید و ردیابی را در دفترکل بلاکچین ثبت می‌کنند. این اطلاعات شامل موارد زیر است:

- هویت فرستنده؛
- شماره دور آموزشی؛
- زمان ثبت اطلاعات؛
- هش ۲۵۶-Keccak مربوط به به‌روزرسانی وزن‌های مدل.

قبل از انجام فرایند تجمیع، کنترل‌کننده می‌تواند بررسی کند که آیا بردار پارامترهای دریافت‌شده خارج از بلاکچین، با هش ثبت‌شده روی بلاکچین مطابقت دارد یا خیر.

این فرایند باعث حفظ موارد زیر می‌شود:

- صحت و یکپارچگی به‌روزرسانی مدل؛
- قابلیت ردیابی تاریخچه تبادل اطلاعات؛
- امکان تشخیص تغییر یا دستکاری پارامترهای مدل.

توابع قرارداد هوشمند شامل موارد زیر خواهند بود:

- ثبت شرکت‌کنندگان، شامل Controller و دستگاه‌ها؛
- بازکردن و بستن دورهای آموزشی؛
- ثبت اطلاعات مربوط به به‌روزرسانی مدل‌ها.

۴. مجموعه داده

مطالعه موردی پیشنهادی، نگهداری و تعمیرات پیش‌بینانه زیرساخت‌های پمپ آب است.

در این مطالعه، متغیرهای عملیاتی به یک برجسب دودویی مربوط به خطر خرابی تبدیل می‌شوند:

- عملکرد عادی؛
- احتمال خرابی قریب‌الوقوع.

انتخاب این مطالعه موردی برای تأیید به استاد راهنما ارائه شده است. در صورت نیاز، می‌توان از هر دیتاست مناسب دیگری برای طبقه‌بندی دودویی استفاده کرد؛ بدون اینکه معماری اصلی سیستم تغییر کند.

دیتاست باید دارای شرایط زیر باشد:

- اندازه مناسب برای پایان‌نامه کارشناسی ارشد؛
- تعداد ویژگی‌های منطقی و قابل مدیریت؛
- مناسب برای طبقه‌بندی دودویی؛
- قابل استفاده با مدل رگرسیون لجستیک.

دو دیتاست عمومی پیشنهاد شده‌اند:

گزینه اول: دیتاست یکپارچه نگهداری و تعمیرات پیش‌بینانه اینترنت اشیا

IoT Integrated Predictive Maintenance Dataset

این دیتاست شامل داده‌های مصنوعی و داده‌های مبتنی بر حسگر است و متغیرهایی مانند موارد زیر را بررسی می‌کند:

- فشار؛
- سرعت؛
- دما.

گزینه دوم: داده‌های حسگر پمپ

Pump Sensor Data

این دیتاست شامل داده‌های فیزیکی مترکم مربوط به موارد زیر است:

- لرزش؛
- دما؛
- اطلاعات حسگرهای یک سامانه شهری.

در ماه اول، پس از انجام تحلیل اکتشافی داده‌ها یا **EDA**، یکی از این دو دیتاست به‌عنوان دیتاست اصلی انتخاب خواهد شد.

در تحلیل اولیه موارد زیر بررسی می‌شوند:

- میزان تعادل یا عدم تعادل کلاس‌ها؛
- داده‌های گمشده؛
- کیفیت داده‌ها؛
- مناسب‌بودن ویژگی‌ها برای مدل رگرسیون لجستیک.

اگر زمان کافی وجود داشته باشد، دیتاست دوم می‌تواند برای بررسی استحکام و پایداری نتایج استفاده شود.

۵. معماری سیستم و فناوری‌های مورد استفاده

نمودار صفحه سوم معماری پیشنهادی سیستم را نشان می‌دهد.

در این معماری:

- یک Controller مدل سراسری را مدیریت می‌کند؛
- سه Device داده‌های محلی جداگانه دارند؛
- مدل سراسری از Controller به دستگاه‌ها ارسال می‌شود؛
- دستگاه‌ها مدل را روی داده‌های محلی خود آموزش می‌دهند؛
- وزن‌های به‌روزشده به Controller بازگردانده می‌شوند؛
- اطلاعات و هش به‌روزرسانی‌ها از طریق تراکنش‌های Metadata روی بلاکچین خصوصی ثبت می‌شوند.

فناوری‌های پیشنهادی عبارت‌اند از:

هماهنگ‌سازی یادگیری فدرال

از چارچوب **Flower** یا **flwr** برای هماهنگی فرایندهای Client-Server در پایتون استفاده می‌شود.

ساختار شامل موارد زیر است:

- یک-Controller ؛
- سه فرایند مربوط به دستگاه‌ها.

یادگیری ماشین

از رگرسیون لجستیک کتابخانه **scikit-learn** به عنوان مدل محلی و مدل سراسری استفاده خواهد شد.

قراردادهای هوشمند

قراردادهای هوشمند با زبان **Solidity** توسعه داده می‌شوند.

برای توسعه، کامپایل و تست قراردادهای می‌توان از یکی از ابزارهای زیر استفاده کرد:

- Hardhat ؛
- Foundry.

قراردادها روی شبکه خصوصی و محلی Geth مستقر خواهند شد.

محیط قرارداد هوشمند

از Hardhat یا Foundry برای موارد زیر استفاده می‌شود:

- توسعه قراردادهای Solidity ؛
- انجام تست واحد؛

- کامپایل خودکار؛
- پیگیری نسخه‌های قرارداد.

گره دفترکل خصوصی

از **Go-Ethereum** یا **Geth** برای اجرای یک نمونه محلی از بلاکچین خصوصی اتریوم استفاده می‌شود.

ارتباط شبکه و اسکرپت‌ها

در صورت نیاز، می‌توان از **Ethers.js** نسخه ۶ در محیط **Node.js** برای ارتباط با قراردادهای هوشمند استفاده کرد.

ارتباط پایتون با بلاکچین

ابزار اصلی پیشنهادی **Web3.py** است.

با استفاده از **Web3.py**، توابع قرارداد هوشمند مستقیماً از داخل حلقه آموزش پایتون فراخوانی می‌شوند.

استفاده از **Ethers.js** و **Node.js** فقط یک گزینه جایگزین است؛ در صورتی که استفاده از یک واسطه جاوااسکریپتی مناسب‌تر باشد.

ابزارهای توسعه

از **Git** برای کنترل نسخه استفاده می‌شود و تمام آزمایش‌ها به‌صورت اسکریپت‌نویسی‌شده اجرا خواهند شد تا نتایج قابل تکرار باشند.

۶. معیارهای ارزیابی

سیستم در چند سناریوی عملیاتی مختلف ارزیابی خواهد شد:

- داده‌های IID در مقایسه با Non-IID؛
- تعداد متفاوت دورهای آموزش؛
- فعال بودن ثبت اطلاعات روی بلاکچین در مقایسه با غیرفعال بودن آن؛
- مقایسه FedAvg و FedProx، در صورت تأیید این توسعه توسط استاد؛
- مقایسه نتایج با مدل متمرکز رگرسیون لجستیک.

معیارهای عملکرد یادگیری فدرال

موارد زیر اندازه‌گیری خواهند شد:

- Accuracy یا دقت کلی؛
- Precision؛
- Recall؛

- F1-score؛
- تغییر عملکرد مدل در دورهای متوالی؛
- نمودار همگرایی Loss مدل سراسری؛
- مقایسه روش‌های تجمیع در صورت استفاده از FedProx.

معیارهای سربار بلاکچین

موارد زیر اندازه‌گیری خواهند شد:

- تأخیر تراکنش یا Transaction Latency؛
- زمان تأیید تراکنش بر حسب میلی‌ثانیه؛
- میزان Gas مصرفی برای هر مرحله ثبت اطلاعات؛
- هزینه ارتباطی در هر دور؛
- میزان افزایش حجم دفترکل بلاکچین.

۷. دستاوردهای مورد انتظار

دستاوردهای مورد انتظار پایان‌نامه عبارت‌اند از:

۱. طراحی و پیاده‌سازی یک سیستم شبیه‌سازی شده BCFL مطابق طرح استاد، شامل:
 - یک-Controller؛
 - سه-Device؛
 - مدل Logistic Regression.
۲. یکپارچه‌سازی بلاکچین خصوصی اتریوم با استفاده از Geth و فرایند یادگیری فدرال از طریق قراردادهای هوشمند.
۳. تحلیل تجربی رفتار یادگیری فدرال در شرایط IID و Non-IID.
۴. در صورت تأیید استاد، مقایسه FedAvg و FedProx.
۵. اندازه‌گیری کمی سربار بلاکچین، شامل:
 - تأخیر؛
 - مصرف Gas؛
 - فضای ذخیره‌سازی.
۶. ایجاد یک نمونه اولیه قابل تکرار و مبتنی بر اسکریپت که بتواند مبنایی برای پژوهش‌های آینده باشد.

۸. برنامه اجرایی شش‌ماهه

ماه اول: مرور منابع و طراحی

اهداف فنی

- مطالعه منابع مربوط به BCFL ، یادگیری فدرال و اتریوم؛
- انجام تحلیل اکتشافی داده‌ها؛
- پیش‌پردازش داده‌ها؛
- نرمال‌سازی ویژگی‌ها؛
- انتخاب دیتاست اصلی؛
- نهایی‌کردن معماری سیستم.

خروجی‌ها

- یادداشت‌های مرور منابع؛
- انتخاب نهایی دیتاست؛
- طراحی نهایی سیستم.

ماه دوم: پیاده‌سازی و آزمایش یادگیری فدرال

اهداف فنی

- پیاده‌سازی سیستم Flower ؛
- ایجاد یک Controller و سه Device ؛
- استفاده از Logistic Regression ؛
- تقسیم داده‌ها به حالت IID و Non-IID ؛
- آزمایش FedAvg ؛
- آزمایش اختیاری FedProx ؛
- ایجاد مدل متمرکز برای مقایسه.

خروجی‌ها

- حلقه پایدار یادگیری فدرال محلی؛
- اسکریپت‌های تقسیم داده؛
- آمار مربوط به تقسیم داده‌ها.

ماه سوم: توسعه قرارداد هوشمند

اهداف فنی

- راه‌اندازی بلاکچین خصوصی Geth ؛
- نوشتن قراردادهای هوشمند؛
- انجام تست واحد قراردادها با Hardhat ؛
- بررسی صحت عملکرد قراردادها.

خروجی

- کد قرارداد هوشمند تأیید شده و در حال اجرا روی بلاکچین خصوصی.
-

ماه چهارم: یکپارچه‌سازی سیستم

اهداف فنی

- اتصال کامل یادگیری فدرال و بلاکچین با استفاده از Web3.py؛
- خودکار سازی ثبت Metadata؛
- بررسی Hash ها؛
- رفع خطاها؛
- پایدار سازی سیستم؛
- ایجاد اسکریپت‌های خودکار برای استخراج Log های بلاکچین؛
- جمع‌آوری معیار های ارزیابی.

خروجی‌ها

- شبیه‌سازی یکپارچه و پایدار؛
 - اسکریپت‌های خودکار جمع‌آوری داده‌ها.
-

ماه پنجم: بنچمارک و ارزیابی

اهداف فنی

- اجرای آزمایش‌ها در تمام سناریوهای تعیین‌شده؛
- جمع‌آوری نتایج کمی؛
- تحلیل نتایج.

خروجی‌ها

- نمودارهای ارزیابی؛
 - نتایج کامل آزمایش‌ها.
-

ماه ششم: نگارش پایان‌نامه

اهداف

- سازمان‌دهی نتایج؛
- جمع‌بندی یافته‌ها؛

- تکمیل متن پایان‌نامه؛
- تکمیل شکل‌ها و نمودارها؛
- تهیه طرح اولیه مقاله علمی.

خروجی‌ها

- نسخه نهایی پایان‌نامه؛
- پیش‌نویس مقاله علمی.

برای کاهش ریسک‌های پیاده‌سازی، یکپارچه‌سازی اولیه یادگیری فدرال و بلاکچین از نیمه دوم ماه سوم آغاز خواهد شد و در ماه چهارم ادامه پیدا خواهد کرد.

نیمه اول ماه چهارم به یکپارچه‌سازی اختصاص داده می‌شود و زمان باقی‌مانده برای موارد زیر در نظر گرفته خواهد شد:

- رفع خطاها؛
- اعتبارسنجی سیستم؛
- بهبود تدریجی عملکرد؛
- آماده‌سازی برای اجرای آزمایش‌ها.

نگارش پایان‌نامه نیز از ماه دوم و به‌صورت موازی آغاز می‌شود. ابتدا فصل‌های پیشینه و طراحی نوشته خواهند شد.

بنابراین، ماه ششم زمان شروع نگارش پایان‌نامه نیست؛ بلکه برای جمع‌بندی، تکمیل و نهایی‌سازی مطالب استفاده خواهد شد.

ریسک‌های اصلی

مشکلات مربوط به تنظیم Geth و قراردادهای هوشمند

راهکار کاهش ریسک:

- شروع با یک شبکه ساده تک‌گره‌ای؛
- نگهداری تنظیمات پروژه در Git.

پیچیدگی یکپارچه‌سازی

راهکار کاهش ریسک:

- انجام یکپارچه‌سازی به‌صورت مرحله‌ای؛
- شروع با ثبت یک تراکنش آزمایشی ساده در هر دور.

نقاط عطف پروژه

ماه	نقطه عطف
M ^۱	تکمیل مرور منابع و تأیید دیتاست و معماری توسط استاد
M ^۲	اجرای نمونه اولیه یادگیری فدرال با FedAvg
M ^۳	راه اندازی بلاکچین خصوصی و تأیید قراردادهای هوشمند
M ^۴	تکمیل و پایدارسازی سیستم یکپارچه BCFL
M ^۵	تکمیل ارزیابی در تمام سناریوها
M ^۶	تحويل پایان نامه و آماده سازی پیش نویس مقاله

۹. مراحل بعدی و پژوهش های آینده

جلسات پیشنهادی بعدی

دو جلسه بعدی به شکل زیر پیشنهاد شده اند:

۱. جلسه اول در حدود ۱۰ روز آینده برای:
 - تأیید انتخاب دیتاست؛
 - تأیید معماری نهایی؛
 - بررسی محدوده ماه اول.
۲. جلسه دوم در پایان ماه دوم برای:
 - بررسی خط لوله فعال یادگیری فدرال؛
 - مشاهده عملکرد نمونه اولیه سیستم.

تاریخ دقیق جلسات از طریق ایمیل تعیین خواهد شد.

مسیر انتشار مقاله

اندازه گیری کمی سربار بلاکچین روی یک نمونه اولیه قابل تکرار می تواند مبنای مناسبی برای انتشار یک مقاله داور شده باشد.

در صورت اضافه شدن مقایسه FedAvg و FedProx، ممکن است دستاورد پژوهشی قوی تر شود.

در ماه پنجم، فهرستی از مجلات یا کنفرانس های مناسب پیشنهاد خواهد شد.

پژوهش های آینده خارج از محدوده پایان نامه

موارد زیر جزو خروجی های الزامی پایان نامه کارشناسی ارشد نیستند و فقط می توانند در پژوهش های آینده بررسی شوند:

- انجام تجمیع مدل ها روی بلاکچین با **On-Chain Aggregation**؛
- استفاده از روش های تجمیع قابل تأیید؛
- استفاده از اثبات های دانش صفر یا **Zero-Knowledge Proofs**.

۱۰. منابع اولیه

منابع اولیه پیشنهادی عبارت‌اند از:

- مقاله «یادگیری فدرال مبتنی بر بلاکچین: یک مرور جامع» نوشته Hu و Wang ، منتشر شده در سال ۲۰۲۱؛ این مقاله توسط استاد راهنما معرفی شده است.
- مقاله McMahan و همکاران درباره یادگیری کارآمد مدل‌های عمیق از داده‌های غیرمتمرکز، منتشر شده در کنفرانس AISTATS سال ۲۰۱۷؛ این مقاله الگوریتم FedAvg را معرفی می‌کند.
- مقاله Li و همکاران درباره بهینه‌سازی فدرال در شبکه‌های ناهمگون، منتشر شده در MLSys سال ۲۰۲۰؛ این مقاله مربوط به FedProx است.
- مقاله زرد اتریوم نوشته Gavin Wood.
- مقاله بیت‌کوین نوشته ساتوشی ناکاموتو در سال ۲۰۰۸.
- مستندات فنی ابزارهای:
 - Geth
 - Flower
 - Hardhat
 - Web3.py
 - Scikit-learn.

این ترجمه، کل محتوای پنج صفحه‌ای فایل را پوشش می‌دهد. مهم‌ترین بخش برای انتخاب فرد فنی، بخش‌های ۳، ۵ و ۶ هستند؛ چون فردی که Federated Learning را بلد است، باید توانایی خود را در Geth، Solidity، Smart Contract، Web3.py، اتصال بلاکچین به حلقه FL و اندازه‌گیری سر بار بلاکچین نشان دهد.