

# طراحی زمان‌بندی و تجمیع برای یادگیری فدرال ناهمزمان از طریق بی‌اسیم شبکه‌ها

چونگ هسوان هو، ژنگ چن و اریک جی. لارسون

پخش مدل، آموزش محلی و تجمیع مدل.

در هر تکرار، فرآیند تجمیع مدل تنها زمانی می‌تواند شروع شود که همه دستگاه‌ها آموزش محلی را به پایان رسانده باشند. بنابراین، مدت زمان یک تکرار به طور جدی توسط کندترین دستگاه محدود می‌شود. [3] این پدیده که معمولاً در روش‌های همزمان FL مشاهده می‌شود، به عنوان مسئله پراکنده شناخته می‌شود. یک رویکرد برای این مسئله، تغییر رویه همزمان به یک رویه ناهمزمان است، یعنی سرور نیازی ندارد منتظر بماند تا همه دستگاه‌ها آموزش محلی را قبل از انجام تجمیع به‌روزرسانی‌ها به پایان برسانند.

در مقالات علمی، چنین چارچوب FL ناهمزمانی در بسیاری از محیط‌های یادگیری عمیق [5]، [4] به کار گرفته شده است. با این حال، FL کاملاً ناهمزمان با به‌روزرسانی متوالی [5]، [4] می‌تواند به دلیل تبادل مکرر مدل، منجر به هزینه‌های بالای ارتباطی شود. از این رو، ما یک چارچوب FL ناهمزمان با تجمیع دورهای پیشنهاد می‌کنیم که اثر سرگردانی را بدون به‌روزرسانی بیش از حد مدل و تبادل اطلاعات بین سرور و دستگاه‌های شرکت‌کننده از بین می‌برد. در مقایسه با سایر کارهای موجود در FL با به‌روزرسانی‌های ناهمزمان [8]، [6] طرح پیشنهادی ما به راحتی قابل پیاده‌سازی است و به مقدار کمی اطلاعات جانبی نیاز دارد.

اصطلاحات شاخص -یادگیری فدرال، آموزش ناهمزمان، شبکه‌های بی‌اسیم، زمان‌بندی، تجمیع

## مقدمه

آموزش مدل‌های یادگیری ماشین (ML) معمولاً به حجم عظیمی از داده‌ها نیاز دارد. امروزه، تعداد روزافزون دستگاه‌های متصل به شبکه، با ارائه مجموعه‌های بزرگی از داده‌ها که می‌توانند برای آموزش مدل مورد استفاده قرار گیرند، به توسعه الگوریتم‌های یادگیری ماشین کمک کرده است. از آنجایی که نگرانی‌های مربوط به حریم خصوصی در جامعه ما حیاتی شده‌اند، استفاده از داده‌های خصوصی از دستگاه‌های کاربر برای آموزش مدل‌های یادگیری ماشین دشوار شده است. بنابراین، یادگیری فدرال (FL) با پردازش اطلاعات روی دستگاه به دلیل مزایای آن در حفظ حریم خصوصی داده‌ها پیشنهاد شده است. FL یک چارچوب یادگیری ماشین مشارکتی است که در آن چندین دستگاه در آموزش یک مدل جهانی مشترک بر اساس داده‌های محلی موجود شرکت می‌کنند. [1] برخلاف معماری یادگیری ماشین متمرکز که در آن کل مجموعه داده‌های آموزشی باید به صورت مرکزی ذخیره شوند، در یک سیستم یادگیری ماشین، فقط پارامترهای مدل بین دستگاه‌های کاربر و یک سرور پارامتر به اشتراک گذاشته می‌شوند. به دلیل ناهمگونی دستگاه‌های شرکت‌کننده، داده‌های آموزشی محلی ممکن است نامتوازن و غیرمستقل و با توزیع یکسان (iid) باشند، که در FL از چارچوب‌های بهینه‌سازی توزیع‌شده مرسوم که در آن‌ها داده‌های همگن و یکپارچه توزیع‌شده فرض می‌شوند، منحرف می‌کند.

میانگین‌گیری فدرال (FedAvg) یکی از معروف‌ترین و پایه‌ترین الگوریتم‌های [2] FL است که یک فرآیند تکراری دارد.

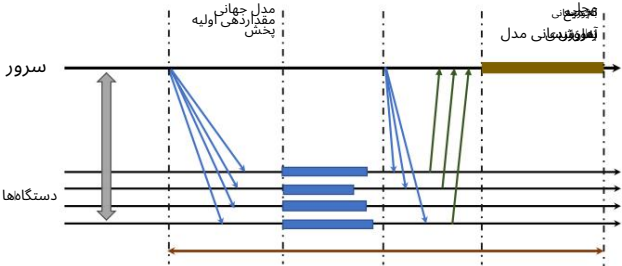
نویسندگان از دپارتمان مهندسی برق (ISY)، دانشگاه لینشوپینگ، 58183 لینشوپینگ، سوئد (ایمیل: chung-hsuan.hu@liu.se, zheng.chen@liu.se, erik.g.larsson@liu.se).

این کار تا حدی توسط Zenith، مرکز تعالی در Linköping - Lund در فناوری اطلاعات (ELLIIT) و بنیاد Knut و Alice Wallenberg پشتیبانی شد.

محدودیت منابع ارتباطی یکی دیگر از مسائل حیاتی در سیستم‌های بی‌اسیم FL است. از آنجایی که تبادل مدل از طریق کانال‌های بی‌اسیم انجام می‌شود، عملکرد سیستم (هزینه‌های ارتباطی و تأخیر) به طور طبیعی از محدودیت منابع فرکانس/زمان رنج می‌برد، به خصوص زمانی که تعداد دستگاه‌های شرکت‌کننده زیاد باشد. یک راه حل ممکن برای کاهش بار ارتباطی، اجازه دادن به بخشی از دستگاه‌های شرکت‌کننده برای آپلود به‌روزرسانی‌های محلی خود برای تجمیع مدل است.

سپس، بسته به منابع ارتباطی اختصاص داده شده و کیفیت لینک بی‌اسیم، هر دستگاه به‌روزرسانی‌های مدل خود را بر این اساس فشرده می‌کند تا به‌روزرسانی‌های فشرده شده با توجه به منابع اختصاص داده شده، به طور قابل اعتمادی منتقل شوند. زمان‌بندی دستگاه و تخصیص منابع ارتباطی برای دستیابی به یادگیری ماشینی کارآمد از نظر ارتباطی در شبکه‌های بی‌اسیم بسیار مهم هستند. [10]، [9] به طور شهودی، دستگاه‌هایی که تأثیر بیشتری بر عملکرد یادگیری دارند باید در زمان‌بندی در اولویت قرار گیرند. این هدف ارتباطی مبتنی بر یادگیری در تضاد با طراحی مرسوم مبتنی بر نرخ است که در شبکه‌های سلولی اتخاذ می‌شود، جایی که هدف دستیابی به کارایی طیفی بالاتر یا توان عملیاتی شبکه است. چندین کار موجود، معیارهای مختلفی را برای نشان دادن اهمیت به‌روزرسانی‌های محلی، مانند هنجار به‌روزرسانی‌های مدل [12]، [11] و سن به‌روزرسانی [13] (AOU) در نظر می‌گیرند. با انگیزه دریافت به‌روزرسانی‌های مدل با افت فشرده‌سازی کمتر، برخی از کارها کیفیت لینک بی‌اسیم را در طراحی زمان‌بندی در نظر می‌گیرند. [17]، [14]، [12] از سوی دیگر، برای در نظر گرفتن توزیع داده‌های غیر iid،

عدم قطعیت توزیع داده‌ها در [15] در نظر گرفته شده است، و در [19]، [18] طراحی زمان‌بندی از اصل اولویت بالاتر به دستگاه‌هایی با تنوع بیشتر در داده‌های محلی خود پیروی می‌کند. برخی از کارها بهینه‌سازی مشترک زمان‌بندی دستگاه و تخصیص منابع را به سمت حداقل تأخیر [20] یا اتلاف تجربی [21]، [22] در نظر می‌گیرند. با این وجود، همه آنها سیستم‌های FL همزمان را در نظر می‌گیرند. تعداد کمی از کارهای موجود، طراحی را در یک محیط ناهمزمان در نظر گرفته‌اند. در [23] زمان‌بندی در FL ناهمزمان بر اساس به حداکثر رساندن مجموع مورد انتظار داده‌های آموزشی با توجه به عدم قطعیت شرایط کانال، ورود داده‌ها و منابع ارتباطی محدود در نظر گرفته شده است. با این حال، تأثیر توزیع داده‌های غیر iid در طراحی زمان‌بندی در نظر گرفته نشده است.



یک دور ارتباطی (تکرار سراسری)

شکل: آفرآیند FL و تبادل اطلاعات بین سرور و دستگاه‌های شرکت‌کننده.

در مقایسه با تنظیمات همزمان، FL ناهمزمان باید با ناهمزمانی به‌روزرسانی‌های مدل محلی مقابله کند، زیرا دستگاه‌های مختلف ممکن است آموزش محلی را بر اساس نسخه‌های مختلف مدل جهانی انجام دهند. برخی از طرح‌های تجمع اکتشافی در مقالات بررسی شده‌اند. در [23] دستگاه‌هایی که از تکرارهای گذشته با شکست‌های انتقال بیشتری مواجه شده‌اند، به‌روزرسانی‌های گرادین بزرگ‌شده را ارسال می‌کنند. در [24] وزن‌های بیشتری به ردیف‌های کندتر در فرآیند تجمع داده می‌شود، زیرا ردیف‌های کندتر کمتر در مدل جهانی مشارکت می‌کنند.

هر دو رویکرد با هدف یکسان‌سازی سهم دستگاه‌های مختلف انجام می‌شوند، اگرچه با داده‌های iid، ممکن است همگرایی کند شود زیرا به‌روزرسانی‌های مدل به‌دست‌آمده از مدل‌های جهانی قدیمی‌تر ممکن است حاوی اطلاعات مفید کمی برای نسخه فعلی باشند.

ما خلاصه‌ای از مهم‌ترین دستاوردهای این اثر را ارائه می‌دهیم:

- ما یک چارچوب FL ناهمزمان با تجمع دوره‌ای پیشنهاد می‌کنیم که در حضور پراکنده‌ها به همگرایی سریع دست می‌یابد و از به‌روزرسانی بیش از حد مدل مانند تنظیمات FL کاملاً ناهمزمان جلوگیری می‌کند.

• ما یک سیاست زمان‌بندی پیشنهاد می‌کنیم که به طور مشترک کیفیت کانال و توزیع داده‌های آموزشی را در نظر می‌گیرد، با هدف کاهش واریانس و بایاس به‌روزرسانی‌های مدل جمعی. اثربخشی روش پیشنهادی هم توسط تحلیل همگرایی نظری و هم نتایج شبیه‌سازی پشتیبانی می‌شود.

- ما یک طرح وزن‌دهی آگاه از سن برای تجمع مدل پیشنهاد می‌کنیم تا اثرات ناهمزمانی به‌روزرسانی را کاهش دهیم. • ما تأثیر فشرده‌سازی به‌روزرسانی، ناهمگونی داده‌ها و ناهمزمانی درون تکرار را در تحلیل همگرایی برجسته می‌کنیم، که همچنین یک انگیزه نظری برای طراحی زمان‌بندی ما فراهم می‌کند.

دوم. مدل سیستم

ما یک سیستم FL با دستگاه‌هایی که در آموزش یک مدل یادگیری جهانی مشترک شرکت می‌کنند، که توسط بردار پارامتری یک بعدی  $R$  پارامتری شده است را در نظر می‌گیریم.  $\{1, \dots, N\}$  به عنوان مجموعه دستگاه در سیستم نشان می‌دهیم. هر دستگاه  $N$  مجموعه‌ای از داده‌های آموزشی محلی  $S$  با اندازه  $|S|$  را در خود جای داده است. فرض کنید  $NS = \bigcup_{i=1}^N S_i$  نشان‌دهنده کل مجموعه داده‌ها در سیستم با اندازه  $|S|$  باشد، که در آن  $S_i \cap S_j = \emptyset$ ،  $S_i \neq \emptyset$ .

دستگاه  $N$  را تنظیم کرد.

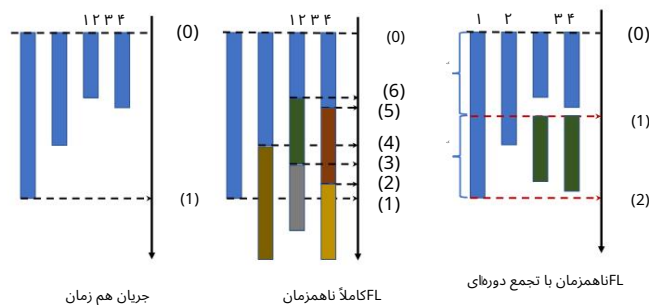
ما مدل  $N$  را به عنوان  $f_N(\cdot)$  نمایش می‌دهیم.

هدف سیستم یافتن بردار پارامتری بهینه است که تابع زیان تجربی تعریف شده توسط را به حداقل برساند. قانون به‌روزرسانی مربوطه به شرح زیر است

(۱) (۲) (۳)

- نشان دهنده شاخص تکرار محلی،  $(\cdot)$  نشان دهنده نرخ یادگیری و  $(0, 0) = (0, 0)$  که در آن  $\eta = 0, \dots$

(۱) 
$$\frac{1}{N} \sum_{i=1}^N f_i(\cdot)$$



عدم تعادل بالقوه مدل ناشی از آموزش ناهمزمان.

ظفاشو محلی منظم در دستگاه در تکرار سراسری-ام و تکرار محلی-ام به صورت زیر تعریف

$$(\cdot) = ((\cdot)) + \mathbb{F} \quad (\cdot) \sqsubset (\cdot, 0)$$

که در آن  $0 < \gamma$  ضریب منظم‌سازی است. برای هر دستگاه  $\Pi$  به‌روزرسانی محلی آن به صورت زیر محاسبه می‌شود

$$+ 1) = ( \cdot ) \square ( \cdot ) \square ( \cdot ) \quad (6)$$

شده اند...،  $0 = w(\cdot, \cdot)$  که روی یک مینی-دسته تصادفی انتخاب شده  $S \cap B(\cdot, \cdot)$  ارزیابی

با توجه به راه اندازی نا همزمان، مدل FL کاملاً نا همزمان و FL همزمان، ( ) نشان دهنده بردار پارامتر مدل در تکرار سراسری -ام است.

$$(\cdot, 0) = (()),$$

کجا

$$= \text{حداکثر}(\{ | \square K(\cdot) \}) + 1$$

آخرین تکرار سراسری را نشان می‌دهد که در آن دستگاه یک مدل سراسری به‌روزرسانی‌شده دریافت کرده است. پس از دریافت به‌روزرسانی‌ها از همه دستگاه‌های موجود در  $\Pi()$ ، سرور تجميع مدل را به صورت زیر انجام می‌دهد:

$$( + 1) = \square \quad ( , 0) + ( \square )$$

$$\square \cap ( )$$

که در آن (در (4) تعریف شده است، و (نشان دهنده وزن  $= 1$ ) است. سپس مدل بازورسانی محلی (محلی) را به عنوان  $W_{local}$  و  $W_{global}$  را به عنوان  $W_{global}$  تعریف می‌کنیم:

$$(\cdot) = \square(\cdot), \quad (7)$$

که نشان دهنده زمان سپری شده از آخرین دریافت یک مدل سراسری به روزرسانی شده است. وزن 2) (می‌تواند به منظور بدیناثری عمل کند(اصولاً، و معمولاً برای این منظور همواره یک نسخه از شبکه ای که انتقال )<sup>(</sup> نشان داده می‌شود. از طریق کانال‌های بی‌سیم منتقل خواهد شد. طرح فشرده‌سازی در زیربخش بعدی به تفصیل شرح داده خواهد شد.

$$(+1) = (+) + \frac{|S| - \max(|S|)}{\max(|S|)} \quad (5)$$

این روش تکراری تا همگرایی ادامه می‌یابد.

ب. FL ناهمزمان با تجمع دوره‌ای

برای پرداختن به مشکل پراکندگی در FL همگام بدون ارباب ارتباطی بیش از حد، ما یک چارچوب FL نامهمزمان با تجمع دوره‌ای پیشنهاد می‌کنیم. ایده کلی این است که آموزش نامهمزمان در دستگاه‌های مختلف امکان‌پذیر شود، به طوری که سرور به صورت دوره‌ای به‌روزرسانی‌ها را از دستگاه‌هایی که محاسبات خود را تکمیل کرده‌اند جمع‌آوری کند، در حالی که بقیه آموزش محلی خود را بدون وقفه یا حذف ادامه می‌دهند. شکل 1 نمونه‌ای از جدول زمانی آموزش و به‌روزرسانی FL همگام اصلی، FL کاملاً نامهمزمان [5]، [4] طرح پیشنهادی ما را نشان می‌دهد. از آنجایی که دستگاه‌های مختلف قابلیت‌های محاسباتی متفاوتی دارند، هر زمان که یک دستگاه آموزش محلی خود را به پایان برساند، سیگنالی به سرور ارسال می‌کند که آمادگی آن را برای گزارش به‌روزرسانی نشان می‌دهد. در هر مدت زمان  $\tau$ ، سرور مجموعه‌ای از دستگاه‌های نامهمزمان به‌روزرسانی را برنامهریزی می‌کند. ما  $K$  را به عنوان مجموع‌های از دستگاه‌های نامهمزمان به‌روزرسانی در یک تکرار ساده  $\tau$ ام تعریف می‌کنیم. فرض کنید  $\Pi(K)$  مجموعه دستگاه‌های برنامهریزی شده باشد، با  $\Pi(K) = \min_i \{K(i)\}$  که در اکثر تعداد دستگاه‌هایی است که می‌توانند برنامهریزی شوند. ما یک عبارت منظم‌سازی در تابع زیان برای کاهش ... در نظر می‌گیریم.

به عنوان

ج. مدل لایه فیزیکی (PHY)

در یک فرایند FL، بخش ارتباط در دو مرحله انجام می‌شود، سرور مدل سراسری را به دستگاه‌ها ارسال می‌کند (انتقال (downlink) و دستگاه‌ها به‌روزرسانی‌های مدل را به سرور گزارش می‌دهند (انتقال (uplink). به‌طور کلی، سرور قدرت انتقال بسیار بالاتری نسبت به دستگاه‌ها دارد و انتقال downlink به دلیل کانال پخش نیازی به تقسیم منابع ارتباطی ندارد. بنابراین، فرض می‌کنیم هیچ خطای فشرده‌سازی در downlink وجود ندارد. با این حال، انتقال uplink گلوگاه ارتباطی رنج می‌برد، که باعث می‌شود طراحی زمان‌بندی و تخصیص منابع از اهمیت ویژه‌ای برخوردار باشد.

2 توجه داشته باشید که این تعریف مبتنی بر سن با تعریف سن به‌روز رسانی (AOU) که در [13] ارائه شده است، متفاوت است، که زمان سپری شده در هر دستگاه را از آخرین مشارکت آن در جمیع عمل اندازه‌گیری می‌کند.

-این مرحله از زمان‌بندی آپ‌لینک به ویژه برای FL از طریق بی‌اسیم اهمیت دارد. شبکه‌ها به عنوان منابع ارتباطی باید بین دستگاه‌ها به اشتراک گذاشته شوند.

پیشینهٔ ریاضیات و کاربرد آن در مهندسی

اندیس‌های  $V()$  جمله دوم  $۳۲$  بیت برای نرم برداری هستند

مقدار، و عبارت سوم به این معنی است که  $1 + (1 + 2^{\log 2})$  بیت هستند

برای انتقال هر عنصر غیر صفر از  $(\wedge)$  مورد نیاز است.

د. انگیزه طراحی زمان‌بندی و تجمیع

تحت تنظیمات FLناهمزمان پیشنهادی، طراحی کلید

سوالات عبارتند از:

(با توجه به محدودیت‌های منابع ارتباطی، چگونه

آیا باید زیرمجموعه‌ای از  $K()$  را برای تجمیع مدل تحت سناریوی داده‌های آموزشی

ناهمگن برنامه‌ریزی کنیم؟

توزیع و کیفیت لینک بی‌اسیم؟

(۲) دستگاه‌های مختلف ممکن است واحدهای محاسبه و منطق (ALU) متفاوتی داشته باشند، یا

جدیدتر یا قدیمی‌تر، چگونه می‌توان یک سیاست وزن‌دهی مناسب با در نظر

گرفتن تازگی طراحی کرد؟

از به‌روزرسانی‌های مدل؟

به یاد داشته باشید که هدف، به حداقل رساندن زیان تجربی جهانی است.

( $\hat{\gamma}$ ) که در (2) داده شده است، ما مدل بهینه را به صورت  $(\argmin)$  تعریف می‌کنیم.

با فرض اینکه  $(\wedge)$  به شدت محدب است، به این معنی که

$0 = (\wedge) = 0$ ، smooth-یعنی برای هر  $R \geq 2$ ،

$$\| \nabla^2 h(\mathbf{x}) \| \leq \frac{1}{\rho} + (2 - \rho) \|\nabla h(\mathbf{x})\| \quad (9)$$

در تکرار  $(+1)$ ام داریم

$$\| \mathbf{x}_t - \mathbf{x}^* \| \leq \frac{1}{\rho} \left( 1 - \frac{\rho}{2} \right)^t \quad (12)$$

که در آن  $E[\cdot]$  نشان‌دهنده‌ی امید به کل تصادفی بودن است

در تکرارهای گذشته  $\dots, 1$ ، توجه داشته باشید که (12) کمیت را تعیین می‌کند

اثربخشی آموزش مدل با بررسی شکاف بین

$$\| \nabla E[\ell(\mathbf{x})] \| \leq \rho \|\nabla \ell(\mathbf{x})\| \quad \text{in } \mathcal{D} \quad (11)$$

(12) را می‌توان به عنوان واریانس مدل تجمیع‌شده تفسیر کرد.

تولید شده از به‌روزرسانی‌های مدل دستگاه‌های برنامه‌ریزی‌شده.

ما برخی از عوامل سیستمی که بر آن تأثیر می‌گذارند را به شرح زیر فهرست می‌کنیم.

(توزیع داده‌های آموزشی: اگر همه دستگاه‌ها iid داشته باشند

با توجه به داده‌های آموزشی، انتظار داریم که  $0 \leq \rho$  باشد، همانطور که در بخش دوم بحث شد.

مستقیماً نتیجه می‌گیرد که

$$\| \nabla^2 \ell(\mathbf{x}) \| \leq \frac{\rho}{\rho} \|\nabla \ell(\mathbf{x})\| \quad (13)$$

تعمیم می‌دهد  $\| \nabla \ell(\mathbf{x}) \| \leq E[\rho] + 1$  است زیرا  $\rho \leq N$  با این حال، با داده‌های آموزشی غیر iid، ممکن است

در دستگاه‌های مختلف زیاد است، که منجر به واریانس زیادی می‌شود

از مدل تجمیع‌شده.

(۲) فشرده‌سازی داده‌ها: همزمان با انتقال به‌روزرسانی‌های مدل

از طریق کانال‌های بی‌اسیم با نرخ محدود، سرور فقط دریافت می‌کند

اطلاعات نویزی به دلیل فشرده‌سازی داده‌ها. تلفات فشرده‌سازی بیشتر، واریانس بیشتری

از مدل تجمیع‌شده ایجاد می‌کند.

(۳) به‌روزرسانی‌های ناهمزمان مدل: همانطور که در بخش II-B نشان داده شده است،

به‌روزرسانی‌های دریافتی  $\nabla \ell(\mathbf{x})$ ، ممکن است LLAهای متفاوتی داشته باشند، که

یک مولف به‌روزرسانی را تغییر در  $(+1)$  خواهد بود -از منظر زمان‌بندی دستگاه، می‌توانیم

کوچکتر  $(+1)$  یا در نظر گرفتن موارد زیر  $\rho \leq 1$

جنبه‌ها:

• دستگاه‌های برنامه‌ریزی‌شده باید یک ساختار همگن ایجاد کنند

نمایش کل مجموعه داده‌های آموزشی S.

ما در نظر می‌گیریم که در هر تکرار سراسری، مدل به‌روزرسانی می‌شود.

از همه دستگاه‌ها از طریق یک کانال محوشدگی بلوکی با بلوک همدوسی نماد  $\theta$  منتقل

می‌شوند. ما متعامد فرض می‌کنیم

$$\nabla \ell(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \nabla \ell(\mathbf{x}_i) \quad (1)$$

بنابراین، انتقال از چندین دستگاه

بدون تداخل. فرض کنید  $h(\cdot)$  نشان دهنده مقیاس بزرگ و

محوشدگی کانال در مقیاس کوچک از دستگاه  $\theta$ -th به

به ترتیب سرور. با فرض اینکه دستگاه  $\theta$ -th ارسال را انجام داده است

توان  $\nabla \ell(\mathbf{x})$  و نویز افزایشی در کانال

دنبال (مقادیر  $\theta$ ) کانال لینک  $\theta$ -ام برابر است با

$$\| \nabla \ell(\mathbf{x}) \| \leq \frac{1}{\rho} + \frac{1}{\rho} \log 2 \quad (8)$$

ما در نظر داریم که دستگاه‌ها فشرده‌سازی داده‌ها و طرح کدگذاری کانال مناسب را مطابق

با کانال اعمال می‌کنند.

ظرفیت در (8) به گونه‌ای که انتقال‌های مدل به‌روزرسانی شوند

بدون خطا هستند. در نتیجه، سرور می‌تواند به طور قابل اعتمادی دریافت کند

تخمین از  $\nabla \ell(\mathbf{x})$  دستگاه.

$\nabla \ell(\mathbf{x})$  و طرح فشرده‌سازی داده‌ها در زیر مشخص شده‌اند.

(اتخصیص منابع ارتباطی: برای دستیابی به

سطح یکسانی از افت فشرده‌سازی در به‌روزرسانی‌های مدل از هر کدام

دستگاه، ما منابع نماد را به گونه‌ای تخصیص می‌دهیم که

$$\nabla \ell(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \nabla \ell(\mathbf{x}_i) \quad (9)$$

که نشان می‌دهد دستگاه‌هایی با کانال‌های بهتر با نمادهای کمتری اختصاص داده

می‌شوند. طراحی مشابه در [11] در نظر گرفته شده است.

(۲) پراکندگی و کمیت‌سازی: هر دستگاه نیاز دارد

فشرده‌سازی  $\nabla \ell(\mathbf{x})$  با توجه به بودجه  $\rho(\cdot)$  بیت‌ها.

ما از پراکندگی تصادفی و به دنبال آن دقت پایین استفاده می‌کنیم.

کوانتیزاسیون تصادفی برای فشرده‌سازی مدل. به طور خاص، ما

یک به‌روزرسانی مدل پراکنده تعریف کنید  $\tilde{\nabla} \ell(\mathbf{x})$

جزء  $1 = \rho$ ، به صورت زیر تعریف می‌شود

$$\tilde{\nabla} \ell(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \nabla \ell(\mathbf{x}_i) \quad (10)$$

در غیر این صورت.

در اینجا،  $\rho(\cdot)$  مؤلفه‌ی  $\rho$ -ام  $\rho(\cdot)$  است و  $V(\cdot)$  نشان دهنده‌ی

مجموعه عناصر رزرو شده با  $V(\cdot) = |\mathcal{V}|$  سپس، هر کدام

عنصر  $\nabla \ell(\mathbf{x})$ ،  $\nabla \ell(\mathbf{x})$  با تصادفی سطح  $\rho$  پردازش می‌شود.

کوانتایزر  $Q: \mathbb{R}^n \rightarrow \mathbb{R}^n$  به طوری که [27]

$$\| Q(\mathbf{x}) - \mathbf{x} \| \leq \frac{1}{\rho} \quad (11)$$

کجا

$$\| Q(\mathbf{x}) - \mathbf{x} \| \leq \frac{1}{\rho} \quad (12)$$

و کوانتیزاسیون است یک متغیر تصادفی است،  $\| Q(\mathbf{x}) - \mathbf{x} \| \leq \frac{1}{\rho}$

$\rho(\cdot)$  دارای ترکیب  $\rho(\cdot)$  است. سطح 3 بردار فشرده حاصل  $\tilde{\nabla} \ell(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \nabla \ell(\mathbf{x}_i)$

این تصادفی کوانتایزر بدون بایاس است، یعنی،  $E[\tilde{\nabla} \ell(\mathbf{x})] = \nabla \ell(\mathbf{x})$  نتنس

$\rho(\cdot)$  و دارای کمیت‌های  $\rho(\cdot) \leq 2$  است.

و  $\rho(\cdot)$  را می‌توان به مقدار  $\rho(\cdot) \leq 2$  محدود کرد.

تحت یک سطح کوانتیزاسیون ثابت، ماکزیمم  $\rho(\cdot)$  را پیدا می‌کنیم.

که قید انتقال را برآورده می‌کند

$$\rho(\cdot) \leq \frac{1}{\rho} + \log 2 + 32 \quad (10)$$

اگر یک بردار صفر باشد، تمام عناصر کوانتیزه شده صفر هستند.

•اتلاف فشرده‌سازی باید پایین نگه داشته شود، که این امر ما را ترغیب می‌کند تا دستگاه‌هایی با شرایط کانال بهتر را در اولویت قرار دهیم.

از منظر تجمع مدل، می‌توانیم با در نظر گرفتن طراحی وزندهی اگاه از سن در فرآیند جمعیت، تأثیر نامطلوب به‌روزرسانی‌های ناهمزمان را کاهش دهیم.

مشکل اعمال به‌روزرسانی‌های قدیمی بر روی مدل تکامل‌یافته. از سوی دیگر، ترجیح به‌روزرسانی‌های محلی جدیدتر به مدل کمک می‌کند تا با گذشت زمان سریع‌تر و روان‌تر همگرا شود، هرچند که این خطر وجود دارد که به یک مدل نامتعادل که به سمت دستگاه‌هایی با قدرت محاسباتی برتر متمایل است، همگرا شود. از آنجایی که سیاست زمان‌بندی پیشنهادی به گونه‌ای انجام می‌شود که همگنی توزیع داده‌ها را بهبود می‌بخشد، مزیت استراتژی «ترجیح مدل‌های جدیدتر» قانع‌کننده‌تر می‌شود.

### III. طراحی زمان‌بندی و تجمیع برای FL ناهم‌زمان

به طور خلاصه، طراحی زمان‌بندی مبتنی بر اهمیت داده و آگاه از کانال با سیاست تجمع ملایم سیاست زمان‌بندی پیشنهاد می‌کند که با در نظر گرفتن توزیع داده‌های آموزشی و شرایط *favoring-fresh* متغیر به عملکرد همگرای سریع و روان می‌آید. برای تأیید اثربخشی طراحی سپس، وزن‌های تجمع در این اساس تنظیم می‌شوند تا تأثیر مضر آموزش نامزمان کاهش یابد. ما، به درجیب تحلیل همگرایی را در بخش ۵.۱ نتایج شبیه‌سازی را در ۱۷ ارائه می‌کنیم.

#### IV. تحلیل همگرایی

ما برخی از نمادگذاری‌ها و تعاریف را برای تبدیل معرفی می‌کنیم.

## تحلیل ماهیت سیستم پیشنهادی.

تعريف ١. (تجميع ناهمزمان):

• فرض کنید (۱) تعداد نسخه‌های مختلف مدل سراسری دریافتی در تجمیع مدل -ام باشد،

یعنی تعداد عناصر منحصر به فرد در مجموعه  $\{() \mid \Pi() \in \Pi()\}$  فرض کنید  $\Pi() = 1, \dots, ()$

$M()$  یک زیرمجموعه دستگاه  $((), M())$  باشد که ALU یکسانی دارد، یعنی  $M()$

همانطور که در سیستم پیشنهادی، زمان‌بندی دستگاه منوط به مجموعه آماده

به‌ارزسانی  $K()$  به جای مجموعه کامل دستگاه  $N$  است، معیارهای زیر را برای تعیین کمیت سطح ناهمگونی داده‌ها برای هر زیرمجموعه دستگاه  $M$  معرفی می‌کنیم.

تعریف ۲. (سطح ناهمگنی داده‌ها): برای یک زیرمجموعه دستگاه  $M \subseteq N$  و تعریف می‌کنیم  $\delta = 1$ ،

$$\Gamma_1(M) = \frac{1}{2} \left( \frac{1}{M} + \frac{1}{M^2} \right)$$

$$\Gamma_2(M) = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in GL_2(\mathbb{C}) \quad (\alpha \neq 0).$$

الف. فرضيات

برای تسهیل تحلیل، فرضیات پایه زیر را در نظر می‌گیریم.

فرض 1. (همواری): تابع زیان محلی  $\ell(\cdot)$ ، همانطور که در (11) مشخص شده است، هموار است که معادل آن است

، به شرط پیوسته لیبشیتز [28] یعنی  $\ell \in C^1(\mathbb{R})$  و  $\ell'(x) \leq L$  برای هر  $x \in \mathbb{R}$ .

$$| \mid \square (1) \square \square (2) \mid | 2 \square \mid | 1 \square 2 \mid | 2;$$

فرض ۲. (تحدب قوی): تابع زیان محلی  $\ell_i(\cdot)$ ، به شدت محدب است، یعنی،  $R$

(1)  $\square$  (2)  $\vdash \square$  (2)  $(\square \vdash \square) \vdash \square$

و به دنبال آن فرضیاتی برای ارزیابی گزادیان تصادفی،

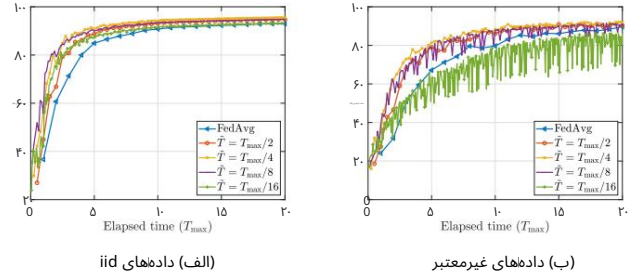
... ، فرض ۳. (محدودیت گشتاور دوم):  $\mathbb{E}[\text{گرادیان تصادفی}] \leq (B; \text{IR})$  ارضا شود

$$\| \mathbf{v}_B \|_2 \leq \| \mathbf{v}_B \|_2 \leq \| \mathbf{v}_B \|_2 \quad (14)$$

$N$ ،  $|S|$  علاوه بر این، دستگاهها در هر  $(\cdot, |S|)$  به حالتی ساده می‌شوند. با  $1 =$

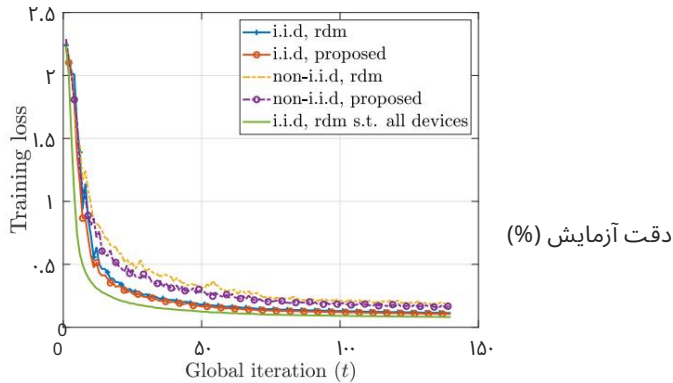
طرح زمان‌بندی ما نه تنها در محیط یادگیری ماشینی ناهمزمان مورد نظر، بلکه برای هر محیط یادگیری ماشینی توزیع‌شده با داده‌های ناهمگن و نامتعادل نیز قابل اجرا است.





سناریوهایی که در آنها  $0.2 = 40\%$  و  $300000 =$  است.

۴- (ناهمزمان) پیشنهادی، و [4] FedAsync توجه داشته باشید که

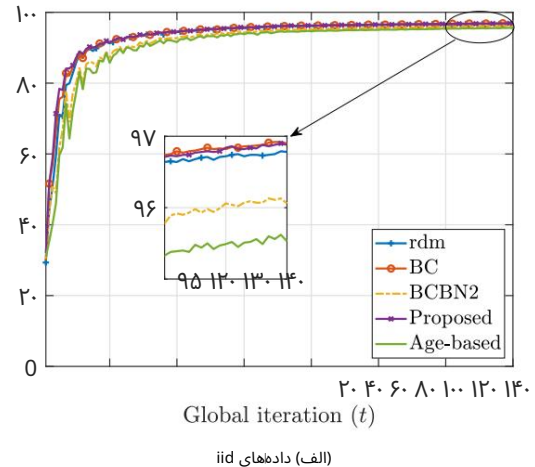


شکل 6: تلفات آموزش برای روش‌های تصادفی و زمان‌بندی پیشنهادی در سناریوهای iid و غیر iid. آن  $\alpha = 0.2$ ,  $\beta = 40$  و  $\gamma = 300000$  است.

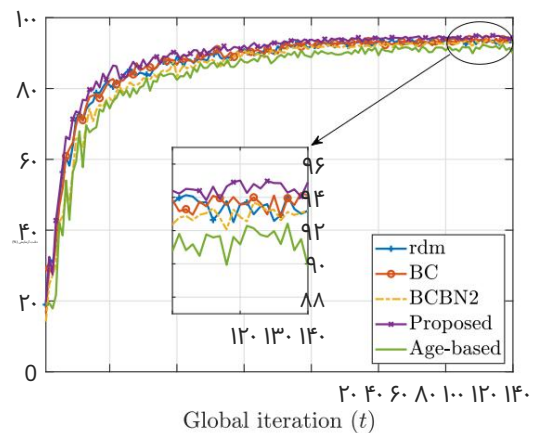
• با روش‌های زمان‌بندی تصادفی و پیشنهادی، افت آموزش در سناریوی iid کمتر از سناریوی غیر iid است، که نشان می‌دهد هرچه کمتر باشد، افت آموزش کمتر است.

• برای هر دو سناریوی iid، non-iid، طرح زمان‌بندی پیشنهادی ما از زمان‌بندی تصادفی بهتر عمل می‌کند، که مزیت انتخاب دستگاه‌هایی با کیفیت لینک بهتر و در مجموع، نمایش داده‌های همگن‌تر را تأیید می‌کند.

• برای نشان دادن تأثیر ناهمزمانی درون تکرار، نتیجه به‌روزرسانی‌های همزمان با زمان‌بندی تصادفی برای داده‌های iid (منحنی سبز) را ارائه می‌دهیم. در مقایسه با حالت ناهمزمان (منحنی آبی)، حالت همزمان تلفات آموزش کمتری دارد. که نشان می‌دهد ناهمزمانی درون تکرار برای عملکرد همگرایی مضر است.



(الف) داده‌های iid



(ب) داده‌های غیرمعتبر

شکل 7: تست دقت سیاست‌های زمان‌بندی مختلف، که در آن  $\alpha = 0.2$ ,  $\beta = 40$  و  $\gamma = 300000$ .

علاوه بر این، همانطور که در شکل 4b با داده‌های غیر iid نشان داده شده است، می‌بینیم که با کوچکتر، که به معنی بزرگتر ( ) و بالاتر است این موضوع، بینش‌های ذکر شده در نکته 4 را تأیید می‌کند که در سناریوهای شدنی داده‌های غیر iid، به‌روزرسانی منجر به تخریب بیشتر عملکرد یادگیری می‌شود.

• مبتنی بر سن 9 [13] ابتدا C را با استفاده از همان روشی که در BCBN2 استفاده شد، پیدا می‌کنیم. سپس، در میان C دستگاه‌هایی را که بالاترین معیار کهنه‌گی را مشخص می‌کنند، به نتیجه دقت 40٪ نسبت به منحنی‌هایی با بزرگترین همگرایی کمتری دارند. این موضوع، بینش‌های ذکر شده در نکته 4 را تأیید می‌کند که در سناریوهای شدنی داده‌های غیر iid، به‌روزرسانی منجر به تخریب بیشتر عملکرد یادگیری می‌شود.

$$\begin{aligned} &= 0 \\ &= 1 - \prod_{i=1}^n (1 - \alpha_i) \end{aligned}$$

( ) = 1 -  $\prod_{i=1}^n (1 - \alpha_i)$  ادر غیر این صورت

می‌بینیم که سیاست زمان‌بندی پیشنهادی، در هر دو سناریوی iid و غیر iid، از زمان‌بندی تصادفی پایه و روش‌های مرجع بهتر عمل می‌کند. علاوه بر این، در سناریوی غیر iid، مشاهده می‌کنیم که آگاهی از داده‌ها در روش پیشنهادی، دقت تست بالاتری نسبت به روش‌های صرفاً آگاه از کانال BC و [11] BCBN2 دارد.

ج. مقایسه سیاست‌های زمان‌بندی

ابتدا، وزن‌های تجمیع با  $\alpha = 1$  در (13) تنظیم می‌شوند. در شکل 7، دقت آزمایش طرح زمان‌بندی پیشنهادی را با برخی روش‌های مرجع جایگزین نشان می‌دهیم.

rdm: ما تا حد اکثر دستگاه‌ها را به طور یکنواخت و تصادفی انتخاب می‌کنیم. BC [11]: ما تا حد اکثر دستگاه‌ها با بالاترین ( ) را انتخاب می‌کنیم. BCBN2 [11]: ابتدا یک زیرمجموعه دستگاه ( ) K C با |C| تا 0.5 پیدا می‌کنیم که بالاترین ( ) دارد. سپس، در میان C، دستگاه‌هایی با بالاترین ( ) را زمان‌بندی می‌کنیم. ( ) | 2.

د. مقایسه سیاست‌های تجمیع

در شکل 8، عملکرد سیاست‌های مختلف تجمیع را تحت روش‌های زمان‌بندی مبتنی بر اهمیت داده و زمان‌بندی تصادفی آگاه از کانال پیشنهادی نشان می‌دهیم. می‌بینیم که طراحی آگاه از سن با  $\alpha = 1$  در از خط پایه بهتر عمل می‌کند.

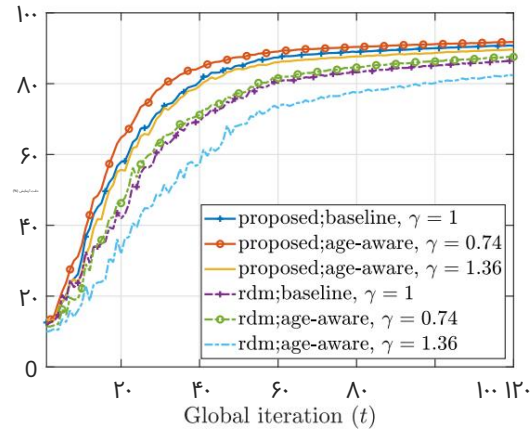
8 برای مقایسه‌ای منصفانه، تخصیص منابع نماد از (9) پیروی می‌کند که متفاوت است از طراحی در [11].

9 این یک نسخه اصلاح‌شده از روش [13] است که به همین ترتیب دستگاه‌هایی با کیفیت کانال بهتر را پیدا می‌کند و میزان کل خرابی دستگاه را به حداقل می‌رساند.

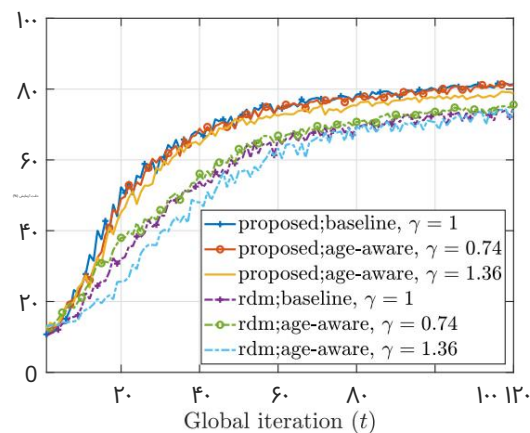


هر دو سناریوی *liid* غیر *liid* برای حالت *liid* استراتژی وزن‌دهی *favoring-fresh* افزایش عملکرد فوق‌العاده‌ای دارد.

برای حالت غیر *liid* افزایش عملکرد طرح جمع‌پیشنهادی زمانی که از زمان‌بندی تصادفی استفاده می‌شود، آشکارتر است.



(الف) داده‌های iid



(ب) داده‌های غیرمعتبر

شکل 8: تست دقت سیاست‌های تجميع پيشنهادی، که در آن  $\alpha = 0.3$ ،  $\beta = 100$ ،  $\gamma = 380000$

### ششم. نتیجه‌گیری

در این کار، ما یک چارچوب FL‌ناهمزمان با تجميع دوره‌ای پیشنهاد کردیم که مزایای آموزش ناهمزمان و تجميع مدل همزمان را با هم ترکیب می‌کند.

برای طرح پیشنهادی، ما یک سیاست زمان‌بندی مبتنی بر اهمیت داده آگاه از کانال و طرح تجمیع مبتنی بر سن برای FL تحت محدودیت‌های منابع پی‌اس‌ام را توسعه دادیم.

طراحی زمان‌بندی و جمع‌پیشنهادی ما، به ویژه با داده‌های آموزشی ناهمگن در بین دستگاه‌های مختلف، عملکرد بهتری نسبت به روش‌های موجود نشان داده است. پیام اصلی این است که اصل طراحی زمان‌بندی و تخصیص منابع در FL با سیم باید مبتنی بر کاهش بایاس و واریانس به‌روزرسانی‌های محلی جمع‌شده باشد. برای تنظیمات FL ناهم‌زمان، متعادل کردن تازگی و مفید بودن به‌روزرسانی‌های مدل محلی و تنظیم سهم آنها در مدل جهانی جدید نیز یک جنبه مهم طراحی است.

هفتم. تقدیر و تشکر

از فردریک جانسون به خاطر مشارکتش در ایده سیاست زمان‌بندی در طول پروژه پایان‌نامه کارشناسی ارشدش در بخش سیستم‌های ارتباطی، دانشگاه لینشوپینگ، تشکر می‌کنیم.

پیوست هشتم

الف. اثبات قضيه ١

ما اثبات را با رویکردی مشابه [25] و [29] انجام می‌دهیم، به جز برخی دستکاری‌های اضافی برای مدیریت تأثیر آموزش ناهمزمان. همانطور که به طور خلاصه در بخش II-D نشان داده شده است، تحلیل شکاف بهیچگی را از (12) شروع می‌کنیم.

تعریف میں آئیں۔

$$(\cdot) = \square \quad (\cdot) \quad \square M(\cdot)$$

به عنوان مجموع وزنی در  $M$  در تکرار سراسری و

$$(\cdot + 1) = \frac{(\cdot)}{M(\cdot)} \quad (\cdot, 1)^{-}$$

و سپس (12) را می‌توان به صورت زیر بازآرایی کرد:

$$E[(+1)] = \frac{1}{r} E[|z|] = \frac{1}{r} \int_0^{\infty} |z| f(z) dz = \frac{1}{r} \int_0^{\infty} z f(z) dz = \frac{1}{r} \int_0^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \frac{1}{r} \left[ -\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \right]_0^{\infty} = \frac{1}{r} \left( \frac{1}{\sqrt{2\pi}} \right) = \frac{1}{r\sqrt{2\pi}}.$$

[illegible]

به ترتیب. علاوه بر این،

$$\varphi(x, 0) = \varphi((x, 0); B(\cdot)).$$

• تخصیص منابع در (9) و محدودیت انتقال در (10) منجر به معادل  $\square$  می‌شود. از این  $\square = \square$

علاوه بر این، ما برخی متغیرها و لم‌ها را معرفی می‌کنیم،

( ) =  $\square$  — ( ) ، ب( ( , 0 ) : [20]

که در آن،  $\tilde{B}(0) = B(0)$ .

$$\bar{E}(\cdot) = E B(\cdot) E Q E V(\cdot) [\bar{E}(\cdot)] = \bar{E}(\cdot) \quad \bar{E}(\cdot) (\cdot, \cdot). \quad (21)$$

۱۴. و مدرا قاضیه کلتیر یاق ( ) شده است، آنگاه

$$\frac{1}{2} \left( \frac{1}{2} + \frac{1}{2} \right) = \frac{1}{2}$$

لم ۲. فرض کنید ۳ و در قضیه تعریف شده باشند، آنگاه واریانس ( ) در رابطه زیر صدق می‌کند:

$$- (1 || 2 || 3) \quad r || - \quad (1) \quad - \quad \frac{1}{r} \quad \frac{r}{-}$$

اثبات به بخش های VIII-B، VIII-Cمراجعه کنید.

می‌تواند به صورت زیر بازنویسی شود  $(+1) \cdot \frac{1}{2}$  بر اساس (20)-(21).

$$\begin{aligned} & \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \\ & \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \end{aligned} \quad (22)$$
$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \quad (23)$$

که در آن (22) به دلیل  $(0, 0) = 0$  است. یا لم 1 و (18) کل انتظار (23) برآورده می‌شود.

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \quad (24)$$

طبق تعریف  $(\cdot)$  و  $E[(\cdot) \cdot (\cdot)] = 0$ . با هم با لم (24)، 2 را می‌توان به صورت زیر بازآرایی کرد:

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

که  $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$  یادگیری  $(\cdot)$  است.  $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$  که شرط کافی در

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \quad (25)$$

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

که  $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$  است.  $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$  که شرط کافی در

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

که در آن (27) به دلیل منفی بودن جمله دوم در (26) است. بنابراین (25) به عبارت سوم (29) برگردانده می‌شود و با جایگذاری نتیجه در (19)

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

که در آن  $EX[\cdot]$  انتظار برای تمام تصمیمات زمان‌بندی گذشته است، یعنی  $\{(\cdot), (\cdot)\}$ ، و  $X = \{M(\cdot) = 1, \dots\}$  مثال  $(\cdot)$  به دلیل (15).

ب. اثبات لم ۱

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

و که در آن (28) و (29) به دلیل تحدب  $2 \cdot \frac{1}{2} = 1$  هستند.  $(\cdot, 0) = (\cdot)$  به ترتیب. عبارت سوم (29) عبارت است از محدود شده توسط تحدب  $(\cdot)$ .

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

با اعمال (30) و (38) در بخش (29)، VIII-D، به صورت 10 به عقب مرتب می‌شود.

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

برای جزئیات به بخش VIII-Dمراجعه کنید.

برابر این تشخیص می‌دهیم

$$\begin{aligned} & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \\ & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \end{aligned}$$

$$\begin{aligned} & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \\ & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \end{aligned}$$

سپس، بر اساس (16) و (17) داریم

$$\begin{aligned} & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \\ & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \end{aligned}$$

توجه داشته باشید که

$$\begin{aligned} & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \\ & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \end{aligned}$$

توجه داشته باشید که

$$\begin{aligned} & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \\ & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \end{aligned}$$

با جایگذاری (34) در (33) داریم

$$\begin{aligned} & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \\ & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \end{aligned}$$

همانطور که حد در (35) به زمان‌بندی فعلی بستگی ندارد سیاست IM، اثبات کامل شده است.

$$\begin{aligned} & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \\ & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \end{aligned}$$

ج. اثبات لم ۲

$$\begin{aligned} & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \\ & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \end{aligned}$$

$$\begin{aligned} & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \\ & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \end{aligned}$$

سمت راست (36) را به عنوان یک تابع جدید تعریف کنید.

$$\begin{aligned} & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \\ & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \end{aligned}$$

از آنجا که (1) یک تابع درجه دوم با حداقل سراسری است ( ) و ( ) نتیجه می‌شود که

$$\begin{aligned} & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \\ & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \end{aligned}$$

سپس، حداقل سراسری برابر است با

$$\begin{aligned} & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \\ & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \end{aligned}$$

با ترکیب (37) و (36)،

$$\begin{aligned} & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \\ & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \end{aligned}$$

و از این رو

$$\begin{aligned} & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \\ & \frac{1}{2} \left( \frac{1}{2} \right)^2 = \frac{1}{4} \end{aligned}$$

[23] ا.اس. لی و جی-دیلیو، لی، «زمان‌بندی انتقال تطبیقی در شبکه‌های بی‌سیم برای یادگیری فدرال ناهمزمان»، مجله IEEE در حوزه‌های منتخب در ارتباطات، جلد 39، شماره 12، صفحات 3673-3687، 2021.

Z. Chai, Y. Chen, A. Anwar, L. Zhao, Y. Cheng, H. Rangwala, «FedAT: یک سیستم یادگیری فدرال با کارایی بالا و کارآمد از نظر ارتباطی با لایه‌های ناهمزمان»، در مجموعه مقالات کنفرانس بین‌المللی محاسبات، شبکه‌سازی، ذخیره‌سازی و تحلیل با کارایی بالا، 2021.

[25] ایکس، لی، کی. هوانگ، دلیو، یانگ، اس. وانگ، و زد، ژانگ، «درباره همگرایی FedAvg روی داده‌های غیر IID» در کنفرانس بین‌المللی بازنمایی‌های یادگیری، 2020.

[26] ای. کوستا، ان. پاپاس، و وی، آنجلایکس، «عصر اطلاعات: یک مفهوم، معیار و ابزار جدید»، مبانی و روندها در شبکه‌سازی، جلد 12، شماره 3، صفحات 162-259، 2017.

[27] دی-آی، ایستار، دی. گروپیک، جی. لی، آر. تومیوکا، و ام. ویتنویچ، «QSGD: SGD با کارایی ارتباطی از طریق کوانتیزاسیون و گذرایی گرادینان»، در مجموعه مقالات سی و یکمین کنفرانس بین‌المللی سیستم‌های پردازش اطلاعات عصبی، 2017، صفحات 1707-1718.

[28] L. Bottou, FE Curtis, Nodedal, «دروش‌های بهینه‌سازی برای یادگیری ماشین در مقیاس بزرگ»، SIAM Review، جلد 60، شماره 2، صفحات 223-311، 2018، [آنلاین]. موجود در: «SGD» SU Stich, [29] doi.org/10.1137/16M1080173 https://openreview.net/ مجله‌ی به سرعت همگرا می‌شود و ارتباط کمی برقرار می‌کند»، در کنفرانس بین‌المللی بازنمایی‌های یادگیری، 2019. [آنلاین].

موجود: forum?id=S1g2JnRcFX [30] LM Nguyen, PH Nguyen, P. Richtárik, K. Scheinberg, M. Takáč, M. van Dijk, «جنبه‌های جدید همگرایی الگوریتم‌های گرادینان تصادفی»، مجله تحقیقات یادگیری ماشین، جلد 20، شماره 176، صفحات 176، 2020.

[31] N. Singh, D. Data, J. George, «مونتیم با کارایی ارتباطی برای بهینه‌سازی غیرمتمرکز»، مجله IEEE در زمینه‌های منتخب در نظریه اطلاعات، جلد 2، شماره 3، صفحات 954-969، 2021.

S. Stich, A. Koloskova, N. Loizou, S. Boreiri, M. Jaggi, «یک نظریه یکپارچه از SGD غیرمتمرکز با تغییر توپولوژی و به‌روزرسانی‌های محلی»، در مجموعه مقالات سی و هفتمین کنفرانس بین‌المللی یادگیری ماشین، جلد 13-18، PMLR، 119، ژوئیه 2020، صفحات 5381-5393.

[33] وای، لکان و سی، کورتس، «پایگاه داده دست‌نویس MNIST»، 2010. [آنلاین]. موجود در: http://yann.lecun.com/exdb/mnist/ [12] ل. بو، دلیو، شیائو، اس. وانگ، جی. هوانگ، و ال، تاسیولاس، «بررسی سیستم و ناهمگونی آماری برای یادگیری فدرال با نمونه‌گیری تطبیقی کلانیت»، در IEEE INFOCOM، 2022، صفحات 1739-1748.

[13] ا.ج اچ یانگ، ای. عرفا، تی کیو کوئک، و اچ وی پور، «سیاست زمان‌بندی مبتنی بر سن برای یادگیری فدرال در شبکه‌های لیه موبایل»، در کنفرانس بین‌المللی IEEE در زمینه آگوستیک، گفتار و پردازش سیگنال، 2020، (ICASSP)، صفحات 8743-8747.

[14] م. صالحی و ای. حسین، «یادگیری فدرال در شبکه‌های بی‌سیم سلولی غیرقابل اعتماد و با محدودیت منابع»، IEEE Trans. on Communications، جلد 69، شماره 8، صفحات 5136-5151، 2021.

[15] دی، لیو، جی، ژو، جی، ژانگ، و کی. هوانگ، «زمان‌بندی کاربر با توجه به اهمیت داده‌ها برای یادگیری ماشینی لیه‌ای با کارایی ارتباطی بالا»، IEEE Trans. در مورد ارتباطات و شبکه‌های شناختی، صفحات 1-1، 2020.

D. Gündüz, ME Ozfatura, J. Zhao, «یادگیری لیه‌ای سریع فدرال با ارتباطات و محاسبات همپوشانی و برنامه‌ریزی منصفانه کلانیت آگاه از کانال»، در بیست و دومین کارگاه بین‌المللی IEEE در مورد پیشرفت‌های پردازش سیگنال در ارتباطات بی‌سیم، 2021، (SPAWC)، صفحات 311-315.

[17] ا.ج. چن، اس. هوانگ، دی. ژانگ، ام. شیائو، ام. اسکاتلند، و اچ. وی. پور، «یادگیری فدرالی بر روی شبکه‌های بی‌سیم اینترنت اشیا با ارتباطات و منابع بهینه»، مجله اینترنت اشیا IEEE، صفحات ۲۰۲۲، ۱-۱.

S. Cherkaoui, A. Taik, H. Moudoud, «زمان‌بندی مبتنی بر کیفیت داده برای یادگیری لیه‌ای فدرال»، در چهل و ششمین کنفرانس IEEE در مورد شبکه‌های کامپیوتری محلی، 2021، (LCN)، صفحات 17-23.

[19] جی. شن، دی. گائو، ال. یانگ، ژو، دی. سونگ، دلیو، لو، و اس. پن، «یادگیری فدرال ناهمگن با کاهش واریانس از طریق انتخاب کلانیت طبقه‌بندی‌شده»، 2022، [آنلاین]. موجود در: [20] https://arxiv.org/abs/2201.05762 «یک چارچوب مشترک دستگاه و تخصیص منابع برای یادگیری فدرال بی‌سیم با محدودیت تأخیر»، IEEE Trans. در ارتباطات بی‌سیم، جلد 20، شماره 1، صفحات 453-467، 2021.

[21] ام. وادو، اس. ساماراکون، و ام. بنیس، «زمان‌بندی مشترک مشتری و تخصیص منابع تحت عدم قطعیت کانال در یادگیری فدرال»، IEEE Trans. در ارتباطات، جلد ۶۹، شماره ۹، صفحات ۵۹۷۴-۵۹۸۲.

[22] ام. چن، زد، یانگ، دلیو، سعد، سی. یین، اچ. وی پور، و اس. کوی، «یک چارچوب مشترک یادگیری و ارتباطات برای یادگیری فدرال از طریق شبکه‌های بی‌سیم»، IEEE Trans. در ارتباطات بی‌سیم، جلد 20، شماره 1، صفحات 269-283، 2021.





تکثیرنگین (عضو IEEE) استانداردها دانشکده مهندسی برق، دانشگاه الیونینگ، سوئد است. او مدرک کارشناسی خود را در سال ۱۳۸۲ از دانشگاه علوم و فناوری هوازیگ (HUST) برای دریافت دکترای خود گرفت. سپس، به ترتیب در سال‌های ۲۰۰۳ تا ۲۰۰۴ مدرک کارشناسی ارشد و دکتری خود را از دانشگاه الیونینگ، سوئد، دانشگاه پاریس ساکل، فرانسه دریافت کرد. از ژانویه ۲۰۱۷ تا اوت ۲۰۱۶، رئیس هیئت مدیره هوشمند تولید شده و شبکه انرژی پیچیده است. او دریافت‌کننده جایزه بهترین مقاله نویسنده جوان انجمن ارتباطات IEEE در سال ۲۰۲۰ بود. او در سال ۲۰۱۶ عنوان Communications Letters، IEEE Communications Transactions و Wireless Transactions on Communications را در سال‌های ۲۰۲۱ و ۲۰۲۲ عنوان رئیس مشترک کارگاه IEEE GLOBECOM در زمینه ارتباطات بیسیم برای هوش مصنوعی خدمت کرد. او در حال حاضر سرپرست Green Communications and Networking در انجمن IEEE است.



اریک جی. لارسون (عضو IEEE) مدرک دکترا را در سال ۲۰۱۲ در دانشگاه ویسلا، اوپسلا، سوئد  
در حال حاضر استاد سیستمهای ارتباطی در دانشگاه لینشوپینگ (LiU) در  
سوئد است، او با موسسه سلطنتی فناوری KTH در استکهلم، سوئد، دانشگاه  
جورج واشینگتن، ایالات متحده آمریکا، دانشگاه فوریاد، ایالات متحده آمریکا و مرکز تحقیقات  
ریگسون، سوئد همکاری داشته است.

علاقه اصلی حرفه‌ای او در حوزه‌های ارتباطات بی‌سیم و پردازش سیگنال است.

او در نگارش کتاب‌های «کدگذاری بلوکی فضا-زمان برای ارتباطات بی‌سیم» (انتشارات دانشگاه کمبریج، ۲۰۰۳ و «مبانی MIMO» (انتشارات دانشگاه کمبریج، ۲۰۱۶ همکاری داشته است.

و به عنوان رئیس کمیته فی الجمله پدیدارن سیگنال (2015-2016) IEEE SPCOM رئیس کمیته راهبردی (2014-2015) Letters  
IEEE Transactions on Wireless-less Communications (2019-2022)، عضو کمیته راهبردی (2019-2022)،  
رئیس عمومی و فی کنفرانس (2012، 2015) Asilomar SSC رئیس مشترک فی کارگاه نظریه ارتباطات (2010) IEEE و عضو هیئت  
معاون اجن پدیدارن سیگنال (2017-2019) IEEE خدمت کرد. از جمله مجلات (2006-2010) on Signal Processing  
IEEE Signal Processing (2010-2014)، IEEE Transactions on Communications (2010-2014)، IEEE Transactions  
بود.

و دو بار در سال‌های ۲۰۱۳ و ۲۰۱۴ جایزه بهترین ستون مجله پردازش سیگنال IEEE در سال ۲۰۱۵ جایزه ComSoc Stephen O. Rice  
IEEE در نظریه ارتباطات، در سال ۲۰۱۷ جایزه ComSoc Leonard G. Abraham IEEE در سال ۲۰۱۸ جایزه بهترین مقاله آموزشی ComSoc  
IEEE در سال ۲۰۱۹ جایزه ComSoc Fred W. Ellersick IEEE دریافت کرد.