

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.

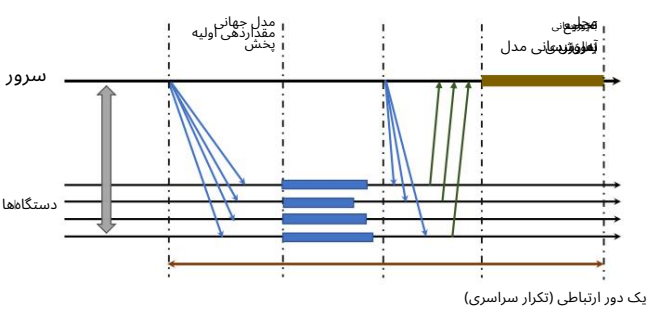
میانگین‌گیری فدرال (FedAvg) یکی از معروف‌ترین و پایه‌ترین الگوریتم‌های [2] FL است که یک فرآیند تکراری دارد.

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.

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.



شکل ۱: فرآیند 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 $S \cap S_i = \emptyset, i \neq 1, \dots, N$.

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)
$$w_{t+1} = w_t - \eta \nabla \ell(w_t; S)$$
 که در آن $(0, 0) = (0, 0)$ نشان دهنده نرخ یادگیری و $\eta = 0, \dots$

•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 تجمیع مدل، می‌توانیم با در نظر گرفتن طراحی وزن‌دهی آگاه از سن در فرآیند تجمیع، تأثیر نامطلوب به‌روزرسانی‌های ناهمزمان را کاهش دهیم.

III.طراحی زمان‌بندی و تجمیع برای FLناهمزمان

به طور خلاصه، طراحی زمان‌بندی مبتنی بر اهمیت داده و آگاه از کانال با سیاست تجمیع *favoring-fresh* منجر به عملکرد همگرایی سریع و روان می‌شود. برای تأیید اثربخشی طراحی سبب، وزن‌های تجمیع بر این اساس تنظیم می‌شوند تا تأثیر مضر آموزش ناهمزمان کاهش یابد. ما؛ به ترتیب تحلیل همگرایی را در بخش IV و نتایج شبیه‌سازی را در V ارائه می‌دهیم.

IV.تحلیل همگرایی

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

تحلیل ماهیت سیستم پیشنهادی.

تعریف ۱.(تجمیع ناهمزمان):

•فرض کنید \mathcal{D} تعداد نسخه‌های مختلف مدل سراسری دریافتی در تجمیع مدل - m باشد،

یعنی تعداد عناصر منحصر به فرد در مجموعه $\mathcal{D} = \{() \mid \Pi() = 1, \dots, M\}$ ،

$M()$ یک زیرمجموعه دستگاه با $() = ()$ ، $M()$ باشد که ALU یکسانی دارد، یعنی $\Pi()$

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

به‌روزرسانی $K()$ 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

$$\Delta(M) = \frac{1}{M} \sum_{i \in M} \Pi(i)$$

$$\Gamma_1(M) = \frac{1}{M} \sum_{i \in M} \Pi(i)$$

$$\Gamma_2(M) = \left(\frac{1}{M} \sum_{i \in M} \Pi(i) \right)^2$$

الف. فرضیات

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

فرض 1.(همواری): تابع زبان محلی $\Pi()$ ، همانطور که در (11) مشخص شده است،

هموار است که معادل آن است

، به شرط پیوسته لپشیتز، [28] یعنی $\Pi \in C^{1,2}$

$$\|\Pi(x) - \Pi(y)\| \leq L \|x - y\|$$

فرض ۲.(تحدب قوی): تابع زبان محلی $\Pi()$ ، (به شدت محدب است، یعنی، R

$$\Pi''(x) \geq \lambda I$$

$$\Pi''(x) \geq (1 - \epsilon) \Pi''(y)$$

و به دنبال آن فرضیاتی برای ارزیابی گرادیان تصادفی،

... فرض ۳.(محدودیت گشتاور دوم): $\Pi()$ گرادیان تصادفی $\Pi(B; \cdot)$ ارضا

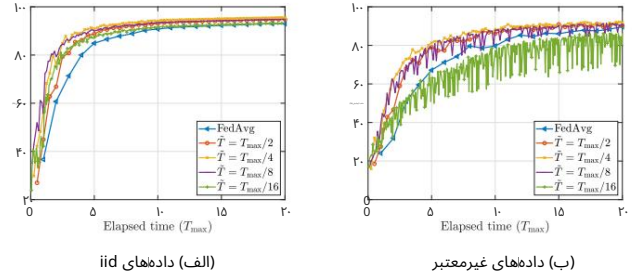
شود

$$\mathbb{E} \|\Pi(B; \cdot) - \Pi(B; \cdot)\|_2^2 \leq \frac{1}{N} \sum_{i=1}^N \mathbb{E} \|\Pi(B; \cdot) - \Pi(B; \cdot)\|_2^2 \quad (14)$$

$$\mathbb{E} \|\Pi(B; \cdot) - \Pi(B; \cdot)\|_2^2 \leq \frac{1}{N} \sum_{i=1}^N \mathbb{E} \|\Pi(B; \cdot) - \Pi(B; \cdot)\|_2^2$$

N . $\mathbb{E} \|\Pi(B; \cdot) - \Pi(B; \cdot)\|_2^2$ به حالتی $\mathbb{E} \|\Pi(B; \cdot) - \Pi(B; \cdot)\|_2^2 = 5$ توجه داشته باشید که $\Gamma_1(N) = \Gamma$

$M()$ دارای ALU یکسانی هستند، بنابراین



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

• (نرخ یادگیری نزولی) $0.01 = (0, 1)$ در ابتدا است
تخاذه شده، همراه با ضریب منظم سازی $0.02 =$

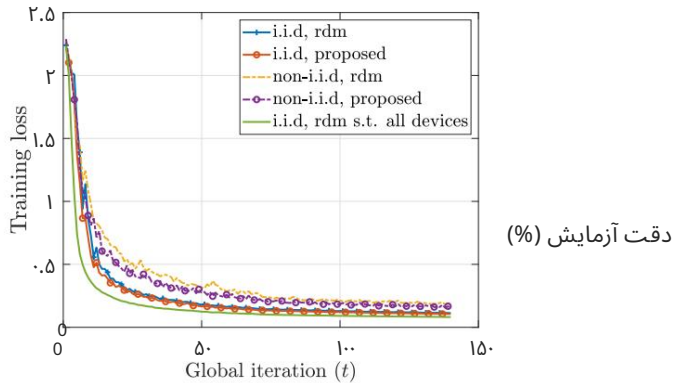
~ در کارهای آینده مورد مطالعه قرار گیرد.

ما فاقد یک عبارت قابل حل هستیم که رابطه بین شکاف بهینگی و مقدار را مشخص کند

(برای مقایسه بین ~های مختلف به زیر مراجعه کنید).

FL ملک ریتہ میگو باجہ اور بھتیجی

۴) **Ft2 (نام برنامه):** پیشنهادی، و [4] FedAsync توجه داشته باشید.



شکل 6: تلفات آموزش برای روش‌های تصادفی و زمان‌بندی پیشنهادی در سناریوهای iid و غیر iid، که در 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 و 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.

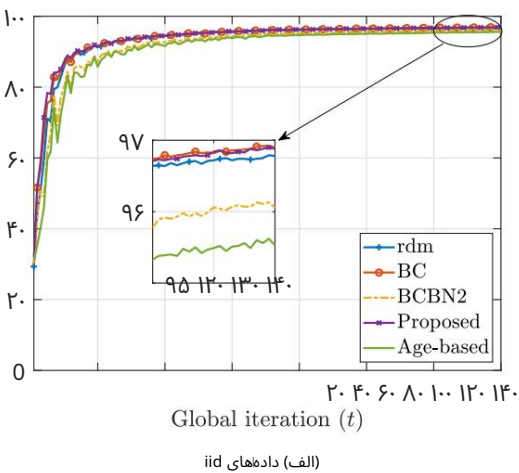
• برای نشان دادن تأثیر ناهمزمانی درون تکرار، نتیجه به‌روزرسانی‌های همزمان با زمان‌بندی تصادفی برای داده‌های iid (منحنی سبز) را ارائه می‌دهیم. در مقایسه با حالت ناهمزمان (منحنی آبی)، حالت همزمان تلفات آموزش کمتری دارد، که نشان می‌دهد ناهمزمانی درون تکرار برای عملکرد همگرایی مضر است.

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 شدن موضوع، بینش‌های ذکر شده در نکته 4 را تأیید می‌کند که در سناریوهای داده‌های غیر iid، درجه بالاتر ناهمزمانی درون تکرار و فشردگی‌سازی به‌روزرسانی منجر به تخریب بیشتر عملکرد یادگیری می‌شود.

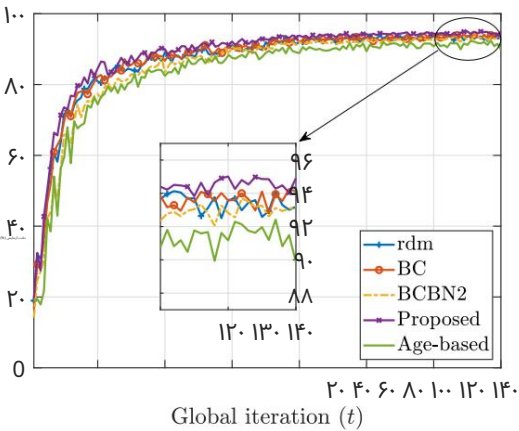
ج. مقایسه سیاست‌های زمان‌بندی

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 حداکثر دستگاه‌ها را به طور یکنواخت و تصادفی انتخاب می‌کنیم.
BC [11]: ما تا حداکثر دستگاه‌ها با بالاترین () را انتخاب می‌کنیم. : 8 [11] BCBN2
• ابتدا یک زیرمجموعه دستگاه () با C | تا 0.5 پیدا می‌کنیم که بالاترین () را دارد. سپس، در میان C، دستگاه‌هایی با بالاترین () را زمان‌بندی می‌کنیم.
() | 2.



(الف) داده‌های iid



(ب) داده‌های غیرمعتبر

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

• مبتنی بر سن : [13]9 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 دارند، زمان‌بندی می‌کنیم. در مقایسه با روش‌های مرجع، پیشنهادی ما بالاترین دقت را در سناریوی iid و غیر iid مشاهده 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.

$$\begin{aligned} &= 0 \\ &(\cdot) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}[\cdot] \end{aligned}$$

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.

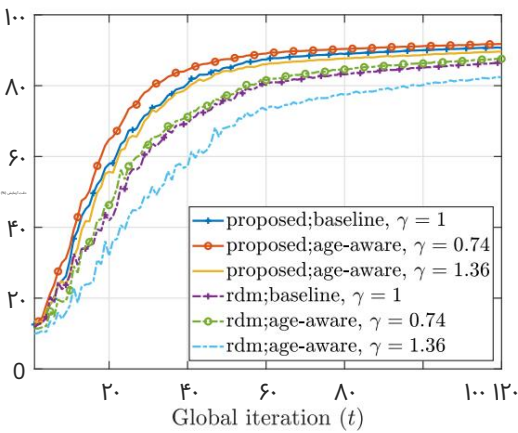
د. مقایسه سیاست‌های تجمیع

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.

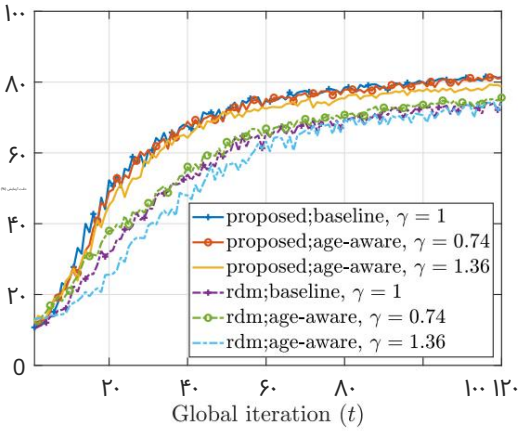
8 برای مقایسه‌ای منصفانه، تخصیص منابع نماد از (9) پیروی می‌کند که متفاوت است از طراحی در [11].

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 iid و غیر iid، برای حالت iid، استراتژی وزندهی favoring-fresh افزایش عملکرد فوق‌العاده‌ای دارد. برای حالت غیر iid، افزایش عملکرد طرح تجمیع پیشنهادی زمانی که از زمان‌بندی تصادفی استفاده می‌شود، آشکارتر است.



(الف) داده‌های iid



(ب) داده‌های غیرمعتبر

شکل 8: تست دقت سیاست‌های تجمیع پیشنهادی، که در آن $\gamma = 0.3, 1, 100, 38000$.

ششم. نتیجه‌گیری

در این کار، ما یک چارچوب FL ناهمزمان با تجمیع دوره‌ای پیشنهاد کردیم که مزایای آموزش ناهمزمان و تجمیع 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 بی‌سیم را توسعه دادیم.

طراحی زمان‌بندی و تجمیع پیشنهادی 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 بی‌سیم باید مبتنی بر کاهش بایاس و واریانس به‌روزرسانی‌های محلی تجمیع‌شده 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.

هشتم. پیوست

الف. اثبات قضیه ۱

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 می‌کنیم.

$$M(\cdot) = \mathbb{E}[\cdot]$$

به عنوان مجموع وزنی در M در تکرار سراسری و

$$\bar{M}(\cdot) = \frac{1}{M} \sum_{i=1}^M M(\cdot)$$

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

$$\bar{M}(\cdot) = \frac{1}{M} \sum_{i=1}^M M(\cdot) = \frac{1}{M} \sum_{i=1}^M \mathbb{E}[\cdot] = \mathbb{E}[\cdot] \quad (19)$$

برای سادگی، ما $\bar{M}(\cdot)$ را به $M(\cdot)$ تغییر می‌دهیم و $\mathbb{E}[\cdot]$ را به $\bar{M}(\cdot)$ تغییر می‌دهیم. برای سادگی، ما $\bar{M}(\cdot)$ را به $M(\cdot)$ تغییر می‌دهیم و $\mathbb{E}[\cdot]$ را به $\bar{M}(\cdot)$ تغییر می‌دهیم.

به ترتیب، علاوه بر این،

$$B(0) = \mathbb{E}[B(0)]$$

*تخصیص منابع در (9) و محدودیت انتقال در (10) منجر به معادل $\bar{M}(\cdot)$ می‌شود. از این رو،

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

$$\bar{M}(\cdot) = \mathbb{E}[\cdot] \quad (20)$$

که در آن $\bar{M}(\cdot) = \mathbb{E}[\cdot]$

$$\bar{M}(\cdot) = \mathbb{E}[\cdot] \quad (21)$$

لم‌ها را فرض کنید $\bar{M}(\cdot)$ شده است، آنگاه

$$\bar{M}(\cdot) = \mathbb{E}[\cdot] \quad (22)$$

لم ۲۰. فرض کنید $\bar{M}(\cdot)$ در قضیه تعریف شده باشند، آنگاه واریانس $\bar{M}(\cdot)$ در رابطه زیر صدق می‌کند:

$$\bar{M}(\cdot) = \mathbb{E}[\cdot] \quad (23)$$

اثبات به بخش های VIII-B، VIII-C مراجعه کنید.

می‌تواند به صورت زیر بازنویسی شود $\| \mathbf{1} + \mathbf{A} \|$ بر اساس (20) ، (21)

[illegible]

که در آن (22) به دلیل \square ، $(, 0) = ()$ است. با لم 1 و

، (18) کل انتظار (23) برآورده می‌شود.

$$\begin{aligned} & \left(\frac{\partial}{\partial t} + \nabla \cdot (\mathbf{v}) \right) \rho = - \rho \nabla \cdot (\mathbf{v}) \\ & \quad + (\rho)_{,t} + (\rho \mathbf{v})_{,i} = - (\rho \mathbf{v})_{,i} \\ & \quad \rho \mathbf{v}_{,i} E_i = (\rho \mathbf{v})_{,i} \end{aligned} \quad ((\rho \mathbf{v})_{,i}). \quad (24)$$

طبق تعريف $w(\cdot)$ ، $E[(\cdot) \square^-(\cdot)] = 0$ ، $\bar{w}(\cdot)$ با هم

با لم (24)، 2 را می‌توان به صورت زیر بازآرایی کرد:

[illegible]

—(+) = بکھکافہاتش تر(+) یادگیری =

(4) $1/(1 + \dots)$ که شرط کافی در

لم ، اما ادعا می‌کنیم که

$$||_9 \quad (i) \quad \square \xrightarrow{+} + \quad (25)$$

۲۰۱۴ که در آن حداکثر $(1) \parallel (+1)$

$$r = \frac{r^2 \left(r^2 + \frac{r^2}{r} \right) + r^2}{\left(\frac{\text{دقیقه}}{1.1(25)10000} \right) r^2}.$$

استقرا ثابت می‌کنیم. برای $n = 1$

$$\frac{1}{(1) \square} = \frac{(1) \square}{(1) \square} = 1$$

کہ (۱) آتہ (۲۶) مرقداں اولیہ باسراستری^۱ ادا مت۔ فرض می کنیم

[illegible]

[illegible]

که در آن (27) به دلیل منفی بودن جمله دوم در (26) است.

بنابراین، (25) برای همه برقرار است و با جایگذاری نتیجه در، (19)

$$\frac{((+1))I_2}{2} \quad () \quad \frac{+}{++1} \quad () \quad \frac{+}{++1}$$

$$[()] \quad \frac{+}{2++1}$$

$$\frac{+}{2} + \frac{+}{2(+1)} \quad (+1) \quad || \quad (1) \quad \frac{+}{2}$$

$$+ \frac{+}{2} + \frac{+}{2} \quad (1) \quad \frac{+}{2} \quad \frac{+}{2}$$

$$(\quad \frac{+}{2} \quad) \quad \frac{+}{2}$$

که در آن $EX[0]$ انتظار برای تمام تصمیمات زمان‌بندی گذشته است، یعنی $\{0, 1\}$

$$_9X = \{M() \mid = 1, \dots,$$

مثال □ [()] به دلیل (15).

ب. اثبات لم ا

[illegible]

و گاه در آن (28) و (29) به دلیل تحدب 2 || . || هستند.

، (0) = () به ترتیب. عبارت سوم (29) عبارت است از

محدود شده توسط تحدب، ()

[illegible]

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

میشود.

[illegible]

برای جزئیات به بخش VIII-D مراجعه کنید.

[23] اچ-اس، لی و جی-دبلیو، لی، «زمان‌بندی انتقال تطبیقی در شبکه‌های بی‌سیم برای یادگیری فدرال ناهمزمان»، مجله IEEE 3673-3687, 2021، صفحات 12، شماره 39، جلد 3.

اچ-اس، لی و جی-دبلیو، لی، «زمان‌بندی انتقال تطبیقی در شبکه‌های بی‌سیم برای یادگیری فدرال ناهمزمان» H. Rangwala، «FedAT: [24] Z. Chai، Y. Chen، A. Anwar، L. Zhao، Y. Cheng، 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، دبلیو، 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، آنجلیکس، «عصر اطلاعات: یک مفهوم، معیار و ابزار جدید»، مبانی و روندها در شبکه‌سازی، جلد 12، شماره 3، صفحات 162-259, 2017.

[27] دی، آلیسترا، دی، گروپیک، جی، لی، آر تومیوکا، و ام، ووینووچ، SGD: QSGD» با کارایی ارتباطی از طریق کوانتیزاسیون و کدگذاری گرادیان»، در مجموعه مقالات سی و یکمین کنفرانس بین‌المللی سیستم‌های پردازش اطلاعات عصبی، 2017، صفحات 1707-1718.

L. Bottou، FE Curtis، [28] Nocedal، «روش‌های بهینه‌سازی برای یادگیری ماشین در مقیاس بزرگ»، SIAM Review، doi.org/10.1137/16M1080173 [29] SU Stich، «SGD موجود در: 223-311 [آنلاین]. موجود در: https://openreview.net/ مجله تحقیقات یادگیری ماشین، جلد 20، شماره 176، صفحات. 2019.

Mوجود: forum?id=51g2JnRCfX [30] LM Nguyen، PH Nguyen، P. Richtárik، K. Scheinberg، M. Takáč، «جنبه‌های جدید همگرایی الگوریتم‌های گرادیان تصادفی»، مجله تحقیقات یادگیری ماشین، جلد 20، شماره 176، صفحات.

[31] N. Singh، D. Data، J. George، jmlr.org/papers/v20/18-759.html موجود در: 149-187, 2019. [آنلاین]. SGD، SquARM-SGD: sgd http:// موجود در: 60، شماره 2، صفحات 223-311. [آنلاین]. موجود در: 954-969, 2021، صفحات 3، شماره 2، جلد 2.

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

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

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

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

[15] دی، لیو، چی، ژو، جی، ژانگ، و کی، هوانگ، «زمان‌بندی کاربر با توجه به اهمیت داده‌ها برای یادگیری ماشینی لبه‌ای با کارایی ارتباطی بالا»، IEEE Trans. 1-1, 2020.

D. Gündüz، ME Ozfatura، J. Zhao، [16] «یادگیری لبه‌ای سریع فدرال با ارتباطات و محاسبات همپوشانی و برنامه‌ریزی منصفانه کلانت آگاه از کانال»، در بیست و دومین کارگاه بین‌المللی IEEE مورد پیشرفت‌های پردازش سیگنال در ارتباطات بی‌سیم ، 2021، SPAWC) صفحات 311-315.

[17] اچ، چن، اس، هوانگ، دی، ژانگ، ام، شیائو، ام، اسکاکلند، و اچ، وی، پور، «یادگیری فدرالی بر روی شبکه‌های بی‌سیم اینترنت اشیا با ارتباطات و منابع بهینه»، مجله اینترنت اشیا ، IEEE، صفحات ۲۰۲۳، ۱-۱.

S. Cherkaoui، [18] A. Taïk، H. Moudoud، «زمان‌بندی مبتنی بر کیفیت داده برای یادگیری لبه‌ای فدرال»، در چهل و ششمین کنفرانس IEEE در مورد شبکه‌های کامپیوتری محلی ، 2021، (LCN) صفحات 17-23.

[19] جی، شن، دی، گاؤ، ال، یانگ، اف، ژو، دی، سونگ، دبلیو، لو، و اس، بن، «یادگیری فدرال ناهمگن با کاهش وابستگی از طریق انتخاب کلانت طبقه‌بندی‌شده»، 2022. [آنلاین]. موجود در: [20] https://arxiv.org/abs/2201.05762، دبلیو، شی، اس، زو، زد، نیو، ام، چیانگ، و ال، گنگ، «زمان‌بندی 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، اوادو، اس، ساماراگون، و ام، بنیس، «زمان‌بندی مشترک مشتری و تخصیص منابع تحت عدم قطعیت کانال در یادگیری فدرال»، IEEE در ارتباطات، جلد ۶۹، شماره ۹۰، صفحات ۵۹۷۴-۵۹۸۲.

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



حاضر سردیبر IEEE Transactions on Green Communications and Networking است.



اریکسون، سوئد همکاری داشته است.

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

کمبریج، ۲۰۰۳ و «مبانی MIMO انبوه» (انتشارات دانشگاه کمبریج، ۲۰۱۶) همکاری داشته است.

IEEE در نظایه ارتباطات، در سال ۲۰۱۷، Leonard G. Abraham، IEEE ComSoc در سال ۲۰۱۸، و مقاله آمینش، ComSoc

IEEE در سال ۲۰۱۹ جایزه IEEE ComSoc Fred W. Ellersick را دریافت کرد.