

تحلیل و مقایسه ۲۱ قید محدودسازی در الگوریتم گرادیان نزولی اتفاقی به روش کرنل

حسین علی غیائی راد^۱ مهدی علیاری شوره دلی^۲ فائزه فریور^۳

۱- دکتری تخصصی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

h.ghiassirad@srbiau.ac.ir

۲- دانشیار - دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران،

aliyari@kntu.ac.ir

۳- استادیار - واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

f.farivar@srbiau.ac.ir

چکیده: مقیدسازی الگوریتم گرادیان نزولی باهدف آموزش شبکه عصبی با وزن‌هایی محدودشده در کاربردهایی نظیر شفاف‌سازی شبکه، کاهش حجم شبکه از نظر ذخیره‌سازی و افزایش سطح عمومیت پذیری آن مؤثر است. همچنین می‌تواند در افزایش سرعت همگرایی، سرعت استنتاج و یافتن جوابی بهینه نیز مناسب باشد. در این نوشتار با استفاده از ترفند کرنل به‌عنوان روشی برای تحمیل انواع قیود بر الگوریتم آموزش، تعداد ۲۱ قید مختلف با یکدیگر مقایسه شده است که تعداد ۱۶ قید آن با الهام از عدم قطعیت موجود در شبکه‌های عصبی زیستی برای اولین بار در این مقاله ارائه شده است. مقایسه قیدها بدون هیچ‌گونه افزایش داده و منظم‌سازی، صورت گرفته است تا اثر قیدها بر فرآیند بهینه‌سازی واضح باشد. به‌منظور ارزیابی، برای هر قید در حل مسائل طبقه‌بندی MNIST، CIFAR-10 و CIFAR-100 با شبکه‌های عصبی عمیق مختص آن، ۶۳ آزمایش شبیه‌سازی شده است. نتایج نشان می‌دهد هر قید در هر مجموعه داده تأثیر متفاوتی بر فرآیند آموزش دارد و به‌طور مشخص قیدهای پیشنهادی که از عدم قطعیت حاضر در شبکه‌های عصبی زیستی الهام گرفته شده است، می‌تواند بهتر از قیدهای ارائه شده در تحقیقات پیشین باشد و باعث بهبود عملکرد شبکه عصبی عمیق از نظر دقت طبقه‌بندی شود.

واژه‌های کلیدی: شبکه عصبی عمیق، شبکه عصبی زیستی، گرادیان نزولی محدودشده، منظم‌سازی، روش کرنل، قید

نوع مقاله: پژوهشی

DOI: 10.52547/jiaeee.19.3.145

تاریخ ارسال مقاله: ۱۳۹۹/۱۲/۰۸

تاریخ پذیرش مشروط مقاله: ۱۴۰۰/۰۷/۲۷

تاریخ پذیرش مقاله: ۱۴۰۰/۱۲/۰۸

نام نویسنده‌ی مسئول: دکتر مهدی علیاری شوره دلی

نشانی نویسنده‌ی مسئول: تهران - خیابان شریعتی - ضلع جنوب پل سیدخندان - دانشگاه صنعتی خواجه‌نصیرالدین طوسی - دانشکده‌ی برق

۱- مقدمه

استفاده از یادگیری محدودشده/مقید به‌ویژه در شبکه‌های عصبی دارای دلایل متعددی است. از جمله این دلایل، آموزش شبکه‌های عصبی تحت قیدهای تحمیلی هنگام پیاده‌سازی سخت‌افزاری است [۱]-[۲]-[۳]-[۴]. مواردی دیگری چون کاهش توان مصرفی [۲]، بالا بردن سرعت محاسبات و دسترسی به حافظه [۲]، و ساده کردن پیاده‌سازی سخت‌افزاری [۱]-[۳] را نیز می‌توان ذکر کرد. انگیزه مهم دیگر در پرداختن به بحث یادگیری تغییرناپذیری الگوها است [۵]. انجام محاسبات چندجمله‌ای‌های دوبعدی [۶]، افزایش سطح عمومیت پذیری و همگرایی سریع‌تر [۷]، نمایش مقاومت به باینری کردن شبکه‌های عصبی عمیق، به هم ریختن و کوانتیزه کردن وزن‌ها [۸]، و همچنین جلوگیری از تداخل و حفظ دانش پیشین نیز از انگیزه‌های دیگر پرداختن به یادگیری محدودشده در شبکه‌های عصبی است.

دو علت و انگیزه دیگر برای پرداختن به یادگیری محدودشده در شبکه‌های عصبی وجود دارد. یکی از این انگیزه‌ها قابل‌فهم کردن و یا شفاف کردن شبکه‌های عصبی است [۹]-[۱۰]، بدین معنی که ویژگی‌هایی که طی آموزش محدودشده در شبکه عصبی عمیق استخراج می‌شود نسبت به زمانی که از قید استفاده نمی‌شود به لحاظ شهودی قابل‌درک‌تر باشند (بینش جدیدی را فراهم کنند). علت و انگیزه دیگر نیز موضوعات زیستی محاسبات نورونی است. این انگیزه باهدف یادگیری بدون ناظر و مدل‌سازی ادراک و شناخت در [۱۱] ارائه شده است. مطالعاتی نیز در خصوص استفاده از قید اشباع برای کنترل تغییرات سیناپسی و جلوگیری از ناپایداری

الگوریتم‌های مبتنی بر STDP^۱ صورت گرفته است [۱۲]. [12]

در این نوشتار با تمرکز بر شبکه‌های عصبی زیستی که ادبیات آن در بخش ۲ مقاله آورده شده است، در جستجوی ابداع قیود تحمیلی جدید و الهام گرفته از زیست‌شناسی است که بتوانند بهبودی در فرآیند آموزش و استنتاج شبکه‌های عصبی عمیق ایجاد کنند. این قیود که با ترفند کرنل [۱۳] به الگوریتم گرادیان نزولی تحمیل می‌شوند در ادامه با قیودی که قبلاً در ادبیات این حوزه ارائه شده است، مقایسه می‌شود. نوآوری‌های ارائه‌شده در این مقاله به‌صورت زیر است:

- توجیه زیستی وجود عدم قطعیت در وزن‌های سیناپسی شبکه عصبی مصنوعی.
- معرفی قیود جدید الهام گرفته از عدم قطعیت موجود در شبکه‌های عصبی زیستی که به‌صورت نویز گوسی به وزن‌های سیناپسی افزوده می‌شود و به بهبود دقت شبکه عصبی عمیق به‌عنوان یک طبقه بند کمک می‌کند.
- معرفی چند قید ابتکاری مانند چرخش وزن‌های لایه‌ها یا قرینه کردن آن‌ها به‌عنوان شکلی از عدم قطعیت.

- مقایسه قیود جدید پیشنهادشده به‌صورت مستقل و ترکیبی با قیود معرفی‌شده در سایر مراجع.

ساختار مقاله بدین‌صورت است: در بخش ۱، ادبیات حوزه یادگیری محدودشده شامل انواع روش‌های تحمیل قیود در زمان آموزش شبکه‌های عصبی مرور شده است. در بخش ۲، مبانی زیستی قیود پیشنهادشده به‌عنوان یکی از نوآوری‌های مقاله تشریح شده است. در بخش ۳، انواع توابع قید معرفی‌شده‌اند. در بخش ۴، آزمایش‌های مختلفی برای مقایسه نتایج اعمال هریک از قیدها طراحی، اجرا و نتایج آن تشریح شده است. نهایتاً در بخش ۵، نتیجه‌گیری بیان‌شده است.

۱-۱- پژوهش‌های پیشین

مقالات متعددی برای معرفی مزایای یادگیری محدودشده/مقید ارائه شده است. محدود کردن تابع فعال‌سازی به‌منظور بهبود پایداری^۲ در آموزش شبکه عصبی مصنوعی از طریق جایگزینی تابع فعال اصلی با یک نمونه محدودشده، از این جمله است. در مقیدسازی پارامترهای شبکه‌های عصبی مصنوعی عمدتاً دو الگوریتم مبتنی بر جستجوی تصادفی و گرادیان نزولی مشاهده می‌شود.

آموزش مقید مبتنی بر جستجوی تصادفی در مواردی چون آموزش قیدهایی که سخت‌افزار بر شبکه تحمیل می‌کند [۱۴] و آموزش شبکه عصبی با وزن‌های عدد صحیح^۳ [۱۵] به کار رفته است. جستجوی تصادفی نیز توسط الگوریتم‌هایی مانند الگوریتم ژنتیک [۱۶] و یا استفاده از توابع تصویر^۴ تصادفی برای مقیدسازی (اشتراک تعداد زیادی از مجموعه‌های مقعر) [۱۷]، انجام شده است.

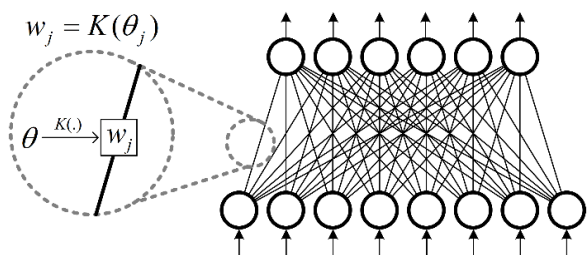
یادگیری محدودشده مبتنی بر گرادیان را نیز می‌توان به دو رویکرد عمده تقسیم کرد. رویکرد اول، روش‌هایی هستند که قید موردنظر را به تابع هزینه یا تابع خطا اضافه می‌کنند. نمونه‌ای از استفاده قید در تابع هزینه در الگوریتم LM^۵ نیز برای آموزش وزن‌های سیناپسی شبکه ارائه‌شده است [۱۸]. نمونه‌ای دیگر، افزودن قیدی به تابع خطا است که در نتیجه آن وزن‌های شبکه عصبی مقادیر غیر منفی به خود گرفته و فیلترهای کانولوشن اسپارس تشکیل می‌دهند [۱۹]. با تغییر تابع هزینه و اضافه کردن مشتق اول تابع فعال که منجر به ظهور مشتق دوم آن در الگوریتم گرادیان نزولی می‌شود، نیز می‌توان قید را اعمال کرد [۲۰]. تخمین نوعی چندجمله‌ای نیز با مقیدسازی تابع هزینه از دیگر کاربردهای این روش است [۲۱].

رویکرد دوم، مقیدسازی مبتنی بر گرادیان از طریق تصویر کردن وزن‌های شبکه عصبی مصنوعی است. در این روش در صورت مشتق‌پذیر بودن تابع تصویر، می‌توان از آن در پس انتشار خطا برای به‌روزرسانی وزن‌ها استفاده کرد. البته نوعی از یادگیری محدودشده مبتنی بر پس انتشار خطا نیز وجود دارد که در آن تابع تصویر مشتق‌پذیر نیست [۲۲]، وزن‌ها توسط تابع ReLU^۶ تصویر می‌شوند و در پس انتشار خطا مشتق آن‌ها محاسبه نمی‌شود. مقالات اغلب از

در این مقاله با الهام از این انواع عدم قطعیت گزارش شده در شبکه‌های عصبی زیستی و با توجه به نوع کاربرد نویز، نویز با توزیع آماری گوسی برای ورود به سیناپس‌ها به‌عنوان نوع جدیدی از قیدها پیشنهاد شده است. همان‌طور که اشاره شد این عدم قطعیت در وزن‌های سیناپسی از نظر زیستی توجیه‌پذیر است.

۳- قیود منتخب به‌عنوان ترفند کرنل^{۱۲}

در ترفند کرنل برای تحمیل قید بر پارامترهای آزاد شبکه، از یک تابع کرنل که قید را فراهم کند، استفاده می‌شود. بنابراین پارامترهای آزاد با کرنل $K(\theta_i)$ جایگزین خواهند شد و سپس الگوریتم آموزشی پارامتر θ_i کرنل را آموزش می‌دهد (شکل (۱)).



شکل (۱): ترفند کرنل برای محدود کردن وزن‌ها

در ادامه لیست قیدهایی که در این مقاله برای آموزش شبکه عصبی مورد استفاده قرار می‌گیرند، ارائه شده است. در مواردی که توضیحات بیشتری نیاز است کد زبان پایتون از کتابخانه Tensorflow [۲۶] در ادامه قید جهت تشریح بیشتر استفاده شده است. در این کدها متغیر w همان Tensor چهاربعدی وزن‌های لایه کانولوشن است و tf نشان‌دهنده کتابخانه Tensorflow است.

- **قیودهای عنصر محور^{۱۳}**. قیودی که بر هر وزن به صورت جداگانه اعمال می‌شوند:
 - ۱- وزن غیر منفی [۱۹] و [۲۲]
 - ۲- وزن محدود: $\|w\|_2 < 2$ (نمونه کرنل: $TanH(\theta_i)$ [۱۳])
 - ۳- وزن‌های نویزی: $w + \gamma$ در صورتی که γ یک متغیر تصادفی با توزیع نرمال، $\gamma \sim N(0, .1)$ است.
 - ۴- وزن‌های نویزی: $w \times \gamma$ در صورتی که γ یک متغیر تصادفی با توزیع $\gamma \sim N(0, .05)$ است.
 - ۵- وزن‌های نویزی: $w(1 + \gamma \times \rho)$ در صورتی که γ یک متغیر تصادفی با توزیع $\gamma \sim N(0, .1)$ و ρ یک ماسک با مقادیر ۰ و ۱ است و باعث می‌شود برخی از وزن‌ها نویزی باشند.
 - ۶- وزن‌های نویزی: در هر فیلتر کانولوشن یک وزن آغشته به نویز وجود دارد. مانند قید شماره ۵، با تنظیم ماسک محقق می‌شود.
 - ۷- وزن‌های نویزی: مانند قید شماره ۶، با این تفاوت که نویز $(\gamma \sim N(0, .2))$ فقط به بخش اول معماری ResNet وارد می‌شود (منظور از بخش در قسمت ساختار آورده شده).

توابع مشتق‌پذیر استفاده کرده‌اند و یا اینکه اگر این توابع مشتق‌پذیر نبوده‌اند، تخمینی مشتق‌پذیر از تابع تصویر را در پس انتشار خطا استفاده کرده‌اند، مانند روش straight through که برای مقیدسازی شبکه‌ای با وزن‌های باینری (۱- و +۱) استفاده شده است [۲۳]. در مواردی نیز از توابعی استفاده شده که در بازه‌ای مشتق‌پذیر بوده‌اند مانند تابع قید شبه خطی^۷ که روی تک‌تک وزن‌های شبکه به صورت جداگانه اعمال شده است [۲۴]. نسخه‌ای از الگوریتم مقید گرادیان نزولی بانام CSGD^۸ نیز ارائه شده است [۲۵] که در آن فرمول به‌روزرسانی وزن‌ها، با در نظر گرفتن تابع تصویری که قید را اعمال می‌کند، اصلاح شده است. این نسخه از الگوریتم گرادیان نزولی از یک تخمین گر بدون سوگیری^۹ از ورودی‌های x ، مقادیر هدف y و در زمان t ، برای تصویر اقلیدسی^{۱۰} استفاده کرده است (به معنی تصویر کردن یک نقطه به یک ابر صفحه). برای اعمال قید L1 نیز الگوریتمی با عنوان SMGD^{۱۱} ارائه شده که در آن با اصلاح معادله به‌روزرسانی گرادیان نزولی برای مقیدسازی، وزن‌ها تصویر می‌شوند [۲۶]. روش‌های مطرح شده به دو علت از عمومیت کمتری نسبت به نسخه اصلی گرادیان نزولی برخوردارند: اولاً، توابع تصویر در آن‌ها مشتق‌پذیر نیست، ثانیاً تغییر در الگوریتم گرادیان نزولی ایجاد کرده‌اند. لذا در این مقاله از روش مبتنی بر گرادیان نزولی با ترفند کرنل برای تحمیل قید استفاده شده است که از عمومیت پذیری بالاتری نسبت به موارد مطرح شده برخوردار است [۱۳].

۲- الهام از مقادیر تصادفی شبکه‌های عصبی

زیستی برای طراحی قیود جدید

وجود مقادیر تصادفی و یا نویز در شبکه‌های عصبی زیستی در مقالات متعددی گزارش شده است. وجود نویز در جمعیت‌های ضربه‌ای که دقت محاسبات نرونی را از آنچه باید باشد کمتر می‌کنند [۲۷]، وجود فواصل زمانی نامنظم بین ضربه‌های تولید شده توسط نرون‌های زیستی و شدت بالای این بی‌نظمی که منجر به تشبیه رفتار نرون در تولید ضربه به انداختن سکه شده است [۲۸] و یا وجود سیگنال‌های تصادفی [۲۹] از این دسته هستند.

وجود نویز در سیناپس‌ها که به صورت آزادسازی تصادفی انتقال‌دهنده‌های عصبی در سیناپس‌های زیستی گزارش شده است [۳۰]، نقطه تمرکز این مقاله است. این رفتار تصادفی در سیناپس‌های زیستی نیز منجر به ایجاد مقادیر نویزی در شبکه‌های عصبی زیستی می‌شود [۳۱].

انواع روش‌های بررسی ماهیت این مقادیر تصادفی یا نویزها در شبکه‌های عصبی زیستی باعث شده تا توزیع‌های آماری مختلفی برای آن‌ها گزارش شود. نویز با توزیع پواسون [۳۲]-[۳۳]-[۳۴] و نویزهایی با توزیع گوسی یا نرمال [۳۵] از مطرح‌ترین انواع توزیع‌های آماری نویز در شبکه‌های عصبی زیستی هستند.

۲۰- قید ترکیبی: استفاده از $\|W_k\|_2 < 1$ با احتمال ۲۰٪، استفاده از γ با احتمال ۲۰٪ و ۶۰ درصد فیلترها نامحدود هستند.

۲۱- قید ترکیبی: استفاده از $\|W_k\|_2 < 1$ با احتمال ۱۰٪، از $\|W_k\|_2 = 1$ با احتمال ۱۰٪، استفاده از γ با احتمال ۱۰٪ و ۷۰ درصد فیلترها نامحدود هستند.

۴- آزمایش

برای ارزیابی تأثیر تحمیل قیدهای مشخص شده در بخش ۳ در سه مسئله طبقه‌بندی CIFAR-10، MNIST و CIFAR-100 تعداد ۶۳ آزمایش انجام شده است. در ادامه معماری شبکه‌های عصبی عمیق مورد استفاده و سپس نتایج آزمایش‌ها گزارش شده است.

۴-۱- ساختارها و تنظیمات

معماری مدل‌های مورد استفاده در آزمایش‌ها، مبتنی بر معماری شناخته شده و محبوب ResNet [۳۸] است و از سه بخش (به‌غیر از لایه کانولوشن ورودی و لایه کاملاً متصل خروجی) تشکیل شده است. در این معماری همانند معماری ResNet، هر بخش متشکل از دو بلوک^{۱۴} روی هم قرار گرفته است. جدول (۱) تعداد فیلترهای هر بخش را نشان می‌دهد. به‌منظور کاهش تأثیرات ناخواسته که مقایسه پذیری نتایج را مختل می‌کند، هیچ‌گونه پیش‌پردازشی بر روی مجموعه داده‌ها (مانند افزایش داده) انجام نشده است.

جدول (۱): تنظیمات معماری ResNet برای آزمایش‌ها. مقدار n_stack در همه بخش‌ها برابر ۲ است

مجموعه داده	تعداد فیلترهای بخش ۱	تعداد فیلترهای بخش ۲	تعداد فیلترهای بخش ۳	تعداد پارامترهای آموزش پذیر
MNIST	۸	۱۶	۳۲	۴۴۸۷۴
CIFAR-10	۳۲	۴۸	۶۴	۲۵۹۱۹۴
CIFAR-100	۴۸	۶۴	۹۶	۶۹۳۹۳۲

۴-۲- نتایج

در این بخش، نتایج اعمال قیدهای ذکر شده در بخش ۳ بر روی مجموعه داده MNIST، CIFAR-10 و CIFAR-100 به ترتیب در

جدول (۲)، جدول (۳) و جدول (۴) ارائه شده است. در این جداول بالاترین دقت طبقه‌بندی فاز تست تحت هر یک از قیدها و ایپاکای^{۱۵} که مدل به آن دقت رسیده است، نشان داده شده تا مشخص شود شبکه عصبی تحت کدام قید و با چه سرعتی به دقت مذکور دست پیدا کرده است.

۸- وزن‌های نویزی: مانند قید شماره ۷، با این تفاوت که نویز فقط به بخش دوم معماری ResNet وارد می‌شود.

۹- وزن‌های نویزی: مانند قید شماره ۷، با این تفاوت که نویز فقط به بخش سوم معماری ResNet وارد می‌شود.

۱۰- وزن‌های نویزی: مانند قید شماره ۶، با این تفاوت که نویز فقط با شدت متفاوت (مقدار میانگین نویز گوسی هرچه به بخش پایانی نزدیک می‌شویم کوچک‌تر می‌شود) در همه بخش‌های معماری ResNet وارد می‌شود. بخش ۱: $\gamma \sim N(0, 2)$ ، بخش ۲: $\gamma \sim N(0, 1)$ ، بخش ۳: $\gamma \sim N(0, 0.05)$.

• **قیدهای در سطح لایه.** در این قیدها، توابع قبلی بر هر وزن جداگانه اعمال می‌شوند و کانولوشن بودن یا تمام متصل بودن لایه تأثیری در آن ندارد. ساختار وزن‌های لایه کانولوشن موردنظر چنین است:

$$rows \times cols \times input_depth \times output_depth$$

• هرکدام از مقادیر از چپ یک محور Tensor (با شروع از صفر) است. برای نمونه در لایه اول کانولوشن شبکه مورد استفاده برای مجموعه داده CIFAR-10 این مقادیر برابر است با:

$$3 \times 3 \times 32 \times 32$$

۱۱- قید شیفت با محور ۲: که در هر مرحله فیلترهای یک‌لایه را یک واحد به راست شیفت می‌دهد (فیلتر سمت راست جایش را با سمت چپ عوض می‌کند).

۱۲- قید معکوس یا فلیپ بر محور ۱: با کد زیر:

$$tf.reverse(w, axis = [1])$$

قید میانگین وزن با فلیپ (معکوس) وزن‌ها بر محور ۳، با کد زیر:

$$(w + tf.reverse(w, axis = [3]))/2$$

۱۴- قید میانگین وزن، با وزن‌های شیفت داده شده، با کد زیر:

$$(w + tf.roll(w, shift = 1, axis = 1))/2$$

۱۵- قید میانگین شیفت چپ و راست با کد زیر:

$$(tf.roll(w, shift = -1, axis = 0) + tf.roll(w, shift = 1, axis = 0))/2$$

• **قیدهای غیر عنصر محور.** قیودی که بر روی تمام وزن‌های یک فیلتر کانولوشن اعمال می‌شوند:

$$16- \text{قیدهای متعامد} (\|W^k\|_2 = 1) \quad [37]$$

$$17- \text{فیلتر کانولوشن محدود} \|W^k\|_2 < 1 \quad [13]$$

$$18- \text{فیلتر کانولوشن محدود} \|W^k\|_2 < 2$$

• **قیدهای ترکیبی.** قیودی که تعدادی از قیدهای عنصر محور و غیر عنصر محور را به‌صورت احتمالاتی بر تعدادی از فیلترهای کانولوشن (بر تمام ماتریس فیلتر اعمال می‌شود) شبکه عمیق اعمال می‌کند:

۱۹- قید ترکیبی: استفاده از $\tanh(w + \gamma)$ با احتمال ۱۵٪ و $\|W_k\|_2 < 1$ با احتمال ۱۵٪ و ۷۰٪ فیلترها بدون قید هستند.

۷	۳۴	۸۴,۳۶	وزن‌های نویزی در بخش ۱	۱۳
۸	۱۱	۸۴,۳۸	وزن‌های نویزی در بخش ۲	۱۰
۹	۱۸	۸۳,۹۸	وزن‌های نویزی در بخش ۳	۱۸
۱۰	۲۹	۸۵,۱۰	وزن‌های نویزی متفاوت در بخش‌ها	۱۰
۱۱	۷۲	۵۱,۹۱	شیفت به راست	۷۲
۱۲	۲۲	۸۲,۳۴	معکوس یا فیلپ	۲۲
۱۳	۱۰	۸۱,۷۱	میانگین وزن با فیلپ آن	۱۰
۱۴	۲۰	۷۹,۶۲	میانگین وزن، با شیفت آن	۲۰
۱۵	۲۳	۷۹,۶۲	میانگین شیفت چپ و راست	۲۳
۱۶	۲۲	۷۶,۵۲	$\ W_k\ _2 = 1$	۲۲
۱۷	۲۴	۸۳,۵۰	$\ W_k\ _2 < 1$	۲۴
۱۸	۲۷	۸۳,۳۳	$\ W_k\ _2 < 2$	۲۷
۱۹	۱۱	۸۴,۱۰	قید ترکیبی	۸
۲۰	۸۹	۸۳,۰۰	قید ترکیبی	۱۱
۲۱	۱۰	۸۳,۶۷	قید ترکیبی	۸۹

جدول (۴): نتیجه آموزش بر روی مجموعه داده CIFAR-100.

نتایجی که از حالت بدون قید بهتر هستند پررنگ شده است

شماره قید	شماره ایپاک	دقت تست (%)	تابع قید	شماره ایپاک
۰	۷	۵۱,۵۳	نامحدود	۷
۱	۱۰۰	۶,۴۴	$0 \leq w$	۱۰۰
۲	۷	۵۱,۹۶	$\ w_i\ < 2$	۷
۳	۳۲	۱۴,۰۱	$w + \gamma$	۳۲
۴	۱۰	۳۱,۸۷	$w \times \gamma$	۱۰
۵	۹	۵۱,۷۹	$w(1 + \gamma \times \rho)$	۹
۶	۱۱	۵۵,۵۰	$w_{1,1} + \gamma$	۱۱
۷	۷	۵۲,۱۳	وزن‌های نویزی در بخش ۱	۷
۸	۱۰	۵۱,۵۴	وزن‌های نویزی در بخش ۲	۱۰
۹	۸	۵۱,۴۷	وزن‌های نویزی در بخش ۳	۸
۱۰	۱۱	۵۲,۱۱	وزن‌های نویزی متفاوت در بخش‌ها	۱۱
۱۱	۹۸	۱۵,۸۹	شیفت به راست	۹۸
۱۲	۱۱	۵۲,۴۵	معکوس یا فیلپ	۱۱
۱۳	۱۰	۴۹,۹۵	میانگین وزن با فیلپ آن	۱۰
۱۴	۱۶	۴۹,۸۰	میانگین وزن، با شیفت آن	۱۶
۱۵	۱۷	۴۸,۷۴	میانگین شیفت چپ و راست	۱۷
۱۶	۵۸	۳۹,۱۱	$\ W_k\ _2 = 1$	۵۸
۱۷	۷	۵۳,۰۲	$\ W_k\ _2 < 1$	۷
۱۸	۸	۵۲,۲۷	$\ W_k\ _2 < 2$	۸
۱۹	۹	۵۲,۶۴	قید ترکیبی	۹

جدول (۲): نتیجه آموزش بر روی مجموعه داده MNIST. نتایجی که از حالت بدون قید بهتر هستند پررنگ شده است

شماره قید	شماره ایپاک	دقت تست (%)	تابع قید	شماره ایپاک
۰	۵	۹۹,۴۴	نامحدود	۵
۱	۹۵	۹۸,۸۰	$0 \leq w$	۹۵
۲	۷	۹۹,۵۰	$\ w_i\ < 2$	۷
۳	۱۰	۶۹,۸۸	$w + \gamma$	۱۰
۴	۲	۹۳,۷۴	$w \times \gamma$	۲
۵	۷	۹۹,۵۳	$w(1 + \gamma \times \rho)$	۷
۶	۲۳	۹۹,۴۴	$w_{1,1} + \gamma$	۲۳
۷	۸	۹۹,۵۵	وزن‌های نویزی در بخش ۱	۸
۸	۳۴	۹۹,۴۱	وزن‌های نویزی در بخش ۲	۳۴
۹	۸	۹۹,۵۷	وزن‌های نویزی در بخش ۳	۸
۱۰	۹	۹۹,۵۰	وزن‌های نویزی متفاوت در بخش‌ها	۹
۱۱	۷۱	۹۷,۹۰	شیفت به راست	۷۱
۱۲	۱۳	۹۹,۴۹	معکوس یا فیلپ	۱۳
۱۳	۲۶	۹۹,۳۰	میانگین وزن با فیلپ آن	۲۶
۱۴	۲۸	۹۹,۳۰	میانگین وزن، با شیفت آن	۲۸
۱۵	۴۴	۹۹,۱۶	میانگین شیفت چپ و راست	۴۴
۱۶	۲۸	۹۹,۳۱	$\ W_k\ _2 = 1$	۲۸
۱۷	۸	۹۹,۵۹	$\ W_k\ _2 < 1$	۸
۱۸	۵	۹۹,۵۳	$\ W_k\ _2 < 2$	۵
۱۹	۱۰	۹۹,۵۳	قید ترکیبی	۱۰
۲۰	۲۴	۹۹,۴۸	قید ترکیبی	۲۴
۲۱	۶	۹۹,۴۵	قید ترکیبی	۶

جدول (۳): نتیجه آموزش بر روی مجموعه داده CIFAR-10.

نتایجی که از حالت بدون قید بهتر هستند پررنگ شده است

شماره قید	شماره ایپاک	دقت تست (%)	تابع قید	شماره ایپاک
۰	۲۸	۸۳,۷۴	نامحدود	۲۸
۱	۹۲	۷۱,۷۳	$0 \leq w$	۹۲
۲	۶	۸۴,۵	$\ w_i\ < 2$	۶
۳	۱۹	۴۸,۱۰	$w + \gamma$	۱۹
۴	۲	۵۷,۱۰	$w \times \gamma$	۲
۵	۱۴	۸۴,۰۷	$w(1 + \gamma \times \rho)$	۱۴
۶	۲۱	۸۳,۹۲	$w_{1,1} + \gamma$	۲۱

۲۰	قید ترکیبی	۵۰.۶۸	۱۱
۲۱	قید ترکیبی	۵۱.۵۹	۹

جدول (۵): جدول کلی موفقیت هریک از قیدها. موفق به معنی دستیابی به دقتی بالاتر از دقت بدون قید است

نتیجه	MNIST	CIFAR-10	CIFAR-100
عدم همگرایی	۱ ۲ ۳ ۴ ۵ ۶ ۷	۱ ۲ ۳ ۴ ۵ ۶ ۷	۱ ۲ ۳ ۴ ۵ ۶ ۷
	۸ ۹ ۱۰ ۱۱ ۱۲ ۱۳ ۱۴	۸ ۹ ۱۰ ۱۱ ۱۲ ۱۳ ۱۴	۸ ۹ ۱۰ ۱۱ ۱۲ ۱۳ ۱۴
	۱۵ ۱۶ ۱۷ ۱۸ ۱۹ ۲۰ ۲۱	۱۵ ۱۶ ۱۷ ۱۸ ۱۹ ۲۰ ۲۱	۱۵ ۱۶ ۱۷ ۱۸ ۱۹ ۲۰ ۲۱
موفق	۱ ۲ ۳ ۴ ۵ ۶ ۷	۱ ۲ ۳ ۴ ۵ ۶ ۷	۱ ۲ ۳ ۴ ۵ ۶ ۷
	۸ ۹ ۱۰ ۱۱ ۱۲ ۱۳ ۱۴	۸ ۹ ۱۰ ۱۱ ۱۲ ۱۳ ۱۴	۸ ۹ ۱۰ ۱۱ ۱۲ ۱۳ ۱۴
	۱۵ ۱۶ ۱۷ ۱۸ ۱۹ ۲۰ ۲۱	۱۵ ۱۶ ۱۷ ۱۸ ۱۹ ۲۰ ۲۱	۱۵ ۱۶ ۱۷ ۱۸ ۱۹ ۲۰ ۲۱
ناموفق	۱ ۲ ۳ ۴ ۵ ۶ ۷	۱ ۲ ۳ ۴ ۵ ۶ ۷	۱ ۲ ۳ ۴ ۵ ۶ ۷
	۸ ۹ ۱۰ ۱۱ ۱۲ ۱۳ ۱۴	۸ ۹ ۱۰ ۱۱ ۱۲ ۱۳ ۱۴	۸ ۹ ۱۰ ۱۱ ۱۲ ۱۳ ۱۴
	۱۵ ۱۶ ۱۷ ۱۸ ۱۹ ۲۰ ۲۱	۱۵ ۱۶ ۱۷ ۱۸ ۱۹ ۲۰ ۲۱	۱۵ ۱۶ ۱۷ ۱۸ ۱۹ ۲۰ ۲۱

به‌طور کلی می‌توان دید که مقیدسازی آموزش وزن‌های سیناپسی با تحمیل عدم قطعیت، از نظر ارتقاء دقت طبقه‌بندی سودمند است. همچنین می‌تواند منجر به کوتاه شدن زمان دستیابی به نتیجه مطلوب شود. با توجه به وجود توجیه زیستی که در بخش ۲ ارائه شد، می‌توان گفت این بهبود ممکن است در اثر ایجاد مقاومت به نویز که در شبکه‌های عصبی زیستی مشاهده می‌شود، رخ داده است. بدین معنی که شبکه یاد می‌گیرد که تغییرات ناشی از چنین عدم قطعیتی کمتر روی عملکرد آن تأثیر بگذارد. از مقایسه نتایج قیود قطعی (ارائه‌شده در مقالات دیگر) و قیود غیرقطعی (ارائه‌شده در این مقاله) می‌توان دریافت که قیود غیرقطعی دقت طبقه‌بندی بالاتری نسبت به قیود قطعی فراهم می‌کنند. یافته‌های حاصل از بررسی دقیق‌تر نتایج گزارش‌شده در جداول این بخش به شرح زیر است:

- جدول (۵) نشان می‌دهد کدام قیدها در کدام مجموعه‌ها موفق یا ناموفق عمل کرده‌اند. در سطر «عدم همگرایی» مشاهده می‌شود که هرچه مجموعه داده سخت‌تر شده تعداد قیدهای ناموفق بالاتر رفته است.
- جدول (۵) نشان می‌دهد در تمام مجموعه داده‌ها به‌غیر از قید شماره ۲ که قبلاً ارائه‌شده، قیدهای ۵، ۶، ۷، ۱۰ و ۱۹ که در این مقاله پیشنهادشده و شامل نوعی عدم قطعیت هستند، نتیجه‌ای بهتر از حالت نامحدود به دست آورده‌اند. این موضوع نشان می‌دهد تزریق عدم قطعیت به وزن‌ها می‌تواند مفید باشد
- جدول (۵)، نتایج قیود شماره ۷، ۸ و ۹ در هر سه مجموعه داده نشان می‌دهد که اولویت افزودن نویز به‌عنوان قید با لایه‌های ابتدایی شبکه است. با در نظر گرفتن قید شماره ۱۰، افزودن نویز به همه لایه‌ها حتی با کاهش بودن حجم نویز از ابتدای شبکه به انتهای آن لزوماً راهبرد مناسبی نیست.

- علاوه بر تزریق عدم قطعیت به وزن‌ها، نتایج قید ۱۹ در هر سه مجموعه داده (جداول ۲، ۳ و ۴) نشان می‌دهد استراتژی تزریق تصادفی قیدها نیز می‌تواند باعث بهبود دقت طبقه‌بندی تصاویر توسط شبکه عصبی عمیق شود، همچنین لزومی ندارد قید حاکم بر آموزش همه وزن‌های سیناپسی همانند باشند.
- با توجه به جداول ۲، ۳ و ۴ در خصوص سرعت دستیابی به بهترین نتیجه (شماره ایپاک) الگوی مشخصی دیده نمی‌شود. در ستون ایپاک آستانه در بیشتر مواقع شماره ایپاک از حالت نامحدود بیشتر است؛ لذا می‌توان گفت با مقیدسازی در صورتی که به نتیجه بهتری برسیم این اتفاق لزوماً سریع‌تر رخ نمی‌دهد.
- نهایتاً پس از بررسی نتایجی که در آن‌ها دقت از حالت نامحدود بسیار پایین‌تر است (مواردی که در جداول ۲، ۳ و ۴ پس‌زمینه شماره خاکستری شده است، ۹ مورد)، مشاهده شد که در این موارد دقت شبکه عصبی در داده آموزشی نیز تقریباً نزدیک داده تست است و اساساً همگرایی رخ نداده، درحالی که در سایر موارد دقت در داده آموزشی تقریباً ۱۰٪ است. پس از حذف این موارد از تمام مواردی که در آن‌ها همگرایی رخ داده میانگین و واریانس گرفته شد (جدول (۶)). نتایج نشان می‌دهد به‌طور میانگین اعمال قید برای مجموعه‌های CIFAR-100 و MNIST عملکرد خوبی داشته.

جدول (۶): میانگین و انحراف معیار نتایج بر روی مواردی که در آن‌ها همگرایی رخ داده

مجموعه داده	دقت در حالت نامحدود	میانگین بدون حالت نامحدود	انحراف معیار بدون حالت نامحدود
MNIST	۹۹.۴۴	۹۹.۰۵	۱.۳۱
CIFAR-10	۸۳.۷۴	۸۲.۱۹	۳.۴۱
CIFAR-100	۵۱.۵۳	۵۱.۷۳	۱.۵۳

۵- نتیجه‌گیری

مقیدسازی آموزش وزن‌های شبکه عصبی توسط ترفند کرنل فرصت مناسبی را برای اعمال انواع قیدها فراهم می‌کند. از طرف دیگر همان‌طور که در مقاله به آن پرداخته شد، زیست‌شناسی و به‌طور خاص شبکه‌های عصبی زیستی مغز انسان می‌تواند منشأ خوبی برای ابداع قیود تحمیلی جدید به فرآیند آموزشی شبکه‌های عصبی عمیق باشد. آنچه از مقایسه عملکرد قیدهای مختلف مشاهده شد نشان داد که اگرچه تأثیر تحمیل قیود می‌تواند تعمیم‌پذیر باشد (رفتار مشابه در هر سه مجموعه داده) اما کاملاً غیرقابل پیش‌بینی است و در واقع نمی‌توان قبل از انجام آزمایش نتیجه مقیدسازی را برآورد کرد. در واقع ما نتوانستیم الگویی را برای طراحی قیدها شناسایی کنیم. اما

- Farivar, "Application of constrained learning in making deep networks more transparent, regularized, and biologically plausible," *Eng. Appl. Artif. Intell.*, vol. 85, 2019, doi: 10.1016/j.engappai.2019.06.022.
- [14] S. S. Liew, M. Khalil-Hani, and R. Bakhteri, "Bounded activation functions for enhanced training stability of deep neural networks on visual pattern recognition problems," *Neurocomputing*, vol. 216, pp. 718–734, 2016, doi: 10.1016/j.neucom.2016.08.037.
- [15] J. R. McDonnell, "Training Neural Networks with Weight Constraints," *Proc. First Annu. Conf. Evol. Program.*, pp. 111–119, 1992.
- [16] Y. Yi, Z. Hangping, and Z. Bin, "A new learning algorithm for neural networks with integer weights and quantized non-linear activation functions," in *IFIP International Federation for Information Processing*, 2008, vol. 276, pp. 427–431.
- [17] M. Wang, Y. Chen, J. Liu, and Y. Gu, "Random Multi-Constraint Projection: Stochastic Gradient Methods for Convex Optimization with Many Constraints," *ArXiv e-prints*, pp. 1–30, 2015, [Online]. Available: <http://arxiv.org/abs/1511.03760>.
- [18] G. Di Muro and S. Ferrari, "A constrained-optimization approach to training neural networks for smooth function approximation and system identification," in *Proceedings of the International Joint Conference on Neural Networks*, 2008, pp. 2353–2359, doi: 10.1109/IJCNN.2008.4634124.
- [19] B. O. Ayinde and J. M. Zurada, "Deep Learning of Constrained Autoencoders for Enhanced Understanding of Data," *IEEE Transactions on Neural Networks and Learning Systems*, no. i, pp. 1–11, 2017.
- [20] F. Han, Q. H. Ling, and D. S. Huang, "Modified constrained learning algorithms incorporating additional functional constraints into neural networks," *Inf. Sci. (Ny.)*, vol. 178, no. 3, pp. 907–919, 2008, doi: 10.1016/j.ins.2007.09.008.
- [21] S. J. Perantonis, N. Ampazis, and S. Varoufakis, "Constrained learning in neural networks: Application to stable factorization of 2-D ...," *Neural Process. ...*, pp. 5–14, 1998, doi: 10.1023/A:1009655902122.
- [22] J. Chorowski and J. M. Zurada, "Learning understandable neural networks with nonnegative weight constraints," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 26, no. 1, pp. 62–69, Jan. 2015, doi: 10.1109/TNNLS.2014.2310059.
- [23] M. Courbariaux, Y. Bengio, and J.-P. David, "BinaryConnect: Training Deep Neural Networks with binary weights during propagations," *CoRR*, Nov. 2015, Accessed: Oct. 23, 2018. [Online]. Available: <http://arxiv.org/abs/1511.00363>.
- [24] A. Bouzerdoum and T. R. Pattison, "Neural Network for Quadratic Optimization with Bound Constraints," *IEEE Trans. Neural Networks*, vol. 4, no. 2, pp. 293–304, 1993, doi: 10.1109/72.207617.
- [25] Y. Mu, W. Ding, T. Zhou, and D. Tao, "Constrained stochastic gradient descent for large-scale least squares problem," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '13*, 2013, p. 883, doi: 10.1145/2487575.2487635.
- [26] A. Tewari, "(Exponentiated) Stochastic Gradient Descent for L1 Constrained Problems," *Learning*, vol. 35900, no. Spring, pp. 1–8, 2008.
- [27] J. Zylberberg, J. Cafaro, M. H. Turner, E. Shea-Brown, and F. Rieke, "Direction-Selective Circuits Shape Noise to Ensure a Precise Population Code,"
 نهایتاً با بررسی میانگین نتایج مقیدسازی می توان توصیه کرد اعمال قید به طور میانگین در برخی مجموعه داده ها (در اینجا MNIST و CIFAR-100) باعث افزایش دقت طبقه بندی می شود. این افزایش دقت در قیود ۲، ۵، ۶، ۷، ۱۰ و ۱۹ مشاهده می شود. همچنین اعمال قیود یا قیدهای غیرقطعی که برای اولین بار در این مقاله ارائه شده است، علاوه بر توجیه زیستی می تواند نتایجی بهتر از قیدهای قطعی موجود از نظر دقت طبقه بندی و حتی سرعت آموزش ایجاد کنند. افزایش دقت آموزش در الگوریتم گرادیان نزولی توسط مقیدسازی می تواند در کاربردهای متنوع شبکه های عصبی [۳۹]–[۴۰] سودمند باشد.

مراجع

- [1] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986, doi: 10.1038/323533a0.
- [2] L. Bottou, "Online learning and stochastic approximations," *On-line Learn. neural networks*, pp. 1–34, 1998.
- [3] J. H. Lee, T. Delbruck, and M. Pfeiffer, "Training deep spiking neural networks using backpropagation," *Front. Neurosci.*, vol. 10, no. NOV, 2016, doi: 10.3389/fnins.2016.00508.
- [4] A. Morro, V. Canals, A. Oliver, M. Alomar, P. Galan-Prado, Fabio Ballester, and J. Rossello, "A Stochastic Spiking Neural Network for Virtual Screening," *IEEE Trans. Neural Networks Learn. Syst.*, p. (In Press), 2017, doi: 10.1109/TNNLS.2017.2657601.
- [5] P. O'Connor and M. Welling, "Deep Spiking Networks," no. Nips, pp. 1–16, 2016, [Online]. Available: <http://arxiv.org/abs/1602.08323>.
- [6] H. Mostafa, "Supervised learning based on temporal coding in spiking neural networks," no. Nips, 2016, [Online]. Available: <http://arxiv.org/abs/1606.08165>.
- [7] S. Sen, S. Venkataramani, and A. Raghunathan, "Approximate computing for spiking neural networks," in *Proceedings of the 2017 Design, Automation and Test in Europe, DATE 2017*, 2017, pp. 193–198, doi: 10.23919/DATE.2017.7926981.
- [8] E. Hunsberger and C. Eliasmith, "Training Spiking Deep Networks for Neuromorphic Hardware," pp. 1–10, 2016, doi: 10.13140/RG.2.2.10967.06566.
- [9] B. Rueckauer, I.-A. Lungu, Y. Hu, and M. Pfeiffer, "Theory and Tools for the Conversion of Analog to Spiking Convolutional Neural Networks," no. Nips, 2016, [Online]. Available: <http://arxiv.org/abs/1612.04052>.
- [10] D. Zambrano and S. M. Bohte, "Fast and Efficient Asynchronous Neural Computation with Adapting Spiking Neural Networks," pp. 1–14, 2016, [Online]. Available: <http://arxiv.org/abs/1609.02053>.
- [11] D. Huh and T. J. Sejnowski, "Gradient Descent for Spiking Neural Networks," *arXiv*, 2017, [Online]. Available: <http://arxiv.org/abs/1706.04698>.
- [12] J. Humble, S. Furber, S. L. Denham, and T. Wennekers, "STDP pattern onset learning depends on background activity," in *Advances in Experimental Medicine and Biology*, 2011, vol. 718, pp. 19–31.
- [13] H. A. Ghiassirad, M. Aliyari Shoorehdeli, and F.

- ⁵ Levenberg–Marquardt
⁶ Rectified Linear Unit
⁷ Semi-linear
⁸ Constrained Stochastic Gradient Descent
⁹ Unbiased estimator
¹⁰ Euclidean projection
¹¹ Stochastic multiplicative gradient descent
¹² Kernel Trick
¹³ Element-wise
¹⁴ Residual block
¹⁵ Epoch

- Neuron, vol. 89, no. 2, pp. 369–383, 2016, doi: 10.1016/j.neuron.2015.11.019.
- [27] W. Maass and W. Maass, “To Spike or Not to Spike: That Is the Question,” Accessed: Sep. 09, 2016. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.702.2268>.
- [29] A. A. Faisal, L. P. J. Selen, and D. M. Wolpert, “Noise in the nervous system,” Nat. Rev. Neurosci., vol. 9, no. april, pp. 292–303, Apr. 2008, doi: 10.1038/nrn2258.
- [30] T. Branco and K. Staras, “The probability of neurotransmitter release: variability and feedback control at single synapses,” Nat. Rev. Neurosci., vol. 10, no. 5, pp. 373–83, May 2009, doi: 10.1038/nrn2634.
- [31] M. D. McDonnell, J. H. Goldwyn, and B. Lindner, “Editorial: Neuronal Stochastic Variability: Influences on Spiking Dynamics and Network Activity,” Front. Comput. Neurosci., vol. 10, p. 38, 2016, doi: 10.3389/fncom.2016.00038.
- [32] D. J. Tolhurst, J. A. Movshon, and A. F. Dean, “The statistical reliability of signals in single neurons in cat and monkey visual cortex,” Vision Res., vol. 23, no. 8, pp. 775–785, 1983, doi: 10.1016/0042-6989(83)90200-6.
- [33] J. M. Beck et al., “Probabilistic Population Codes for Bayesian Decision Making,” Neuron, vol. 60, no. 6, pp. 1142–1152, 2008, doi: 10.1016/j.neuron.2008.09.021.
- [34] R. Moreno-Bote, “Poisson-Like Spiking in Circuits with Probabilistic Synapses,” PLoS Comput. Biol., vol. 10, no. 7, 2014, doi: 10.1371/journal.pcbi.1003522.
- [35] C. Gulcehre, M. Moczulski, M. Denil, and Y. Bengio, “Noisy Activation Functions,” 2016, doi: 10.1063/1.2245436.
- [36] M. Abadi et al., “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.” Accessed: Sep. 06, 2019. [Online]. Available: www.tensorflow.org.
- [37] Q. Le, J. Ngiam, Z. Chen, D. H. Chia, and P. Koh, “Tiled convolutional neural networks,” Nips, pp. 1–9, 2010, [Online]. Available: <https://papers.nips.cc/paper/4136-tiled-convolutional-neural-networks.pdf>.
- [38] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2016, doi: 10.1109/CVPR.2016.90.

[39] حکم آبادی، ع.، ذاکری، آ.، & شرافت، ا. (۲۰۱۷). کاهش سطح

گلببرگ‌های فرعی در فشرده‌سازی پالس رادار برای آشکارسازی اهداف همسان با استفاده از شبکه عصبی. نشریه مهندسی برق و الکترونیک ایران، ۱۴(۲)، ۶۵–۷۳. <http://jiaeee.com/article-1-389-fa.html>

[40] عمادالاسلامی، م.، مجیدی، ح.، & حقی فام، م. (۲۰۲۲). ارائه یک مدل دومرحله‌ای جهت تشخیص تقلب در شبکه توزیع به‌وسیله یادگیری عمیق. نشریه مهندسی برق و الکترونیک ایران، ۱۹(۱). https://jiaeee.com/browse.php?a_id=1319

زیر نویس ها

- ¹ Spike time dependent plasticity
² Constraint Stability
³ Integer
⁴ Projection