

بسم تعالی



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده مهندسی کامپیوتر

شبکه‌های عصبی ضربه‌ای

نگارنده:

محمد رضا مولوی

استاد درس:

رضا صفابخش

تابستان ۱۴۰۱

چکیده

وقتی صحبت از شبکه‌های عصبی مصنوعی (ش.ع.م)¹ می‌شود اولین توضیح ایده گرفتن این شبکه‌ها از مغز انسان است ولی واقعیت این است که نورون‌هایی که به عنوان واحد اصلی محاسبات در این شبکه استفاده می‌شود خیلی با نورون‌های زیستی که در مغز وجود دارند متفاوت است. شبکه‌های عصبی ضربه‌ای (ش.ع.ض)² تلاشی برای مدل سازی آنچه که واقعا در سطح نورونی مغز اتفاق می‌افتد، هستند.

نورون‌های ش.ع.م دارای تابع فعالیت³ پیوسته هستند و به صورت تکی مدل می‌شوند اما نورون‌های زیستی از ضربه‌های گسسته برای انتقال اطلاعات استفاده می‌کنند. به همین علت، ش.ع.ض هم در مکان و زمان خلوت هستند. این امر باعث شده است که ساخت و استفاده از سخت‌افزارهای توان پایین به همراه روش‌های یادگیری محلی ممکن شود. البته آموزش ش.ع.ض همچنان یک مسئله چالش برانگیز است. تابع تحریک این شبکه‌ها معمولا مشتق‌ناپذیر بوده و استفاده از روش‌های آموزش مبتنی بر مشتق مانند روش پس انتشارخطا ممکن نیست.

در پردازش داده‌های حجیم⁴ و وظایفی که نیاز به یادگیری برخط است استفاده از ش.ع.م و دیگر روش‌های دسته‌بندی نیاز به الگوریتم‌هایی دارد که به کمک آن‌ها بتوان تغییرات سریع داده‌ها در حجم بالا را کنترل و رسیدگی کرد. به طور کلی این مدل‌ها اگر در فضاهای متغیر قرار بگیرند نیاز به آموزش مجدد دارند. در حالی که ش.ع.ض خود را به عنوان یکی از موفق‌ترین رویکردها در انجام وظایف یادگیری برخط نشان داده‌اند.

در این پژوهش، به بررسی رویکردهای مختلف استفاده از ش.ع.ض در انواع یادگیری و مقایسه آن‌ها از نقطه نظرات دقت و پیچیدگی خواهیم پرداخت. یک نمای کلی از این پژوهش این است که ش.ع.ض هنوز از ش.ع.م دقت پایین‌تری اما این اختلاف روز به روز در حال کاهش است.

کلمات کلیدی: شبکه‌های ضربه‌ای-یادگیری عمیق-یادگیری آنلاین-یادگیری بانظارت با نورون‌های ضربه‌ای

¹ Artificial Neural Networks

² Spiking Neural Networks

³ Activation Function

⁴ Big Data

فهرست مطالب

- ۱ - مقدمه ۱
- ۲ - شبکه‌های عصبی ضربه‌ای _ ساختار و آموزش ۳
 - ۱-۲ ساختار شبکه‌های ضربه‌ای ۳
 - ۲-۲ قوانین یادگیری در شبکه‌های عصبی ضربه‌ای ۴
 - ۱-۲-۲ یادگیری با نظارت در شبکه‌های عصبی ضربه‌ای ۴
 - ۲-۲-۲ یادگیری بی نظارت در شبکه‌های عصبی ضربه‌ای ۶
 - ۳- یادگیری عمیق در شبکه‌های عصبی ضربه‌ای ۸
 - ۱-۳ شبکه‌های کاملاً متصل ضربه‌ای ۹
- ۴ - شبکه‌های عصبی پیچشی ضربه‌ای ۱۱
- ۵ - شبکه‌های عصبی بازگشتی ضربه‌ای ۱۳
- ۶ - یادگیری بر خط در شبکه‌های عصبی ضربه‌ای ۱۴
- ۷ - نتیجه‌گیری ۱۷

فهرست جداول و تصاویر

- شکل ۱) ساختار یک شبکه عمیق ساده ۸
- شکل ۲) رفتار شبکه با پتانسیل غشائی به عنوان یک سیگنال مشتق پذیر ۱۰
- شکل ۳) آزمایش‌های مختلف برای مقدار پارمترها و دقت حاصل شده برای شبکه کاملاً متصل ۱۱
- شکل ۴) ساختار یک شبکه ضربه‌ای پیچشی ۱۲
- شکل ۵) ساختار یک شبکه ضربه‌ای پیچشی تبدیل شده ۱۳
- جدول ۱) بررسی مدل‌های ضربه‌ای توسعه داده شده برای یادگیری برخط ۱۴

۱- مقدمه

در ش.ع.م واحدهای محاسباتی که به نورون نیز معروف هستند خروجی پیوسته تولید می کنند. این واحدها غالباً مشتق پذیر بوده و از توابع تحریک غیرخطی استفاده می کنند. مشتق پذیری استفاده از روش های مبتنی بر گرادینان را ممکن می سازد و غیرخطی بودن نشان دهنده قدرت این شبکه ها در حل مسائل پیچیده است. موفقیت های شبکه های عمیق در زمینه هایی مثل پردازش زبان طبیعی [2]، پردازش تصویر [1] استفاده از این شبکه ها در صنعت و کاربرد را بسیار رایج کرده است.

ایده شبکه های عمیق الهام گرفته از مغز است، ولی در اصل تفاوت هایی بین واحدهای ش.ع.م با نورون های واقعی وجود دارد. از مهم ترین تفاوت به نحوه انتقال اطاعات بین واحدهای شبکه می توان اشاره کرد. انتشار اطلاعات بین نورون های مغز توسط توالی پتانسیل های الکتریکی (ضربه) صورت می گیرد. این ضربه ها که در زمان خلوت هستند مقدار زیادی اطلاعات را شامل می شوند. هر ضربه شامل یک مقدار انرژی بازدارنده¹ یا وادارنده² است و هر نورون بعد از شلیک شدن یک مقدار تاخیر برای پذیرش ضربه های جدید دارد. پس اطلاعات با زمانبندی این ضربه ها، انرژی و مقدار تاخیر گسترش پیدا می کند.

نورون های ش.ع.م با استفاده از تحریک های پیوسته انتقال اطلاعات را انجام می دهند. این یعنی ش.ع.م فرصتی برای کاهش محاسبات را به وجود می آورند. ضربه ها در زمان خلوت هستند. این شبکه ها به ویژگی های زمانی اطلاعات حساس اند. طبق مطالعاتی که بر روی مغز صورت گرفته هر یک از ضربه ها برای قسمت های ویژه ای از مغز معتبر هستند، که در نتیجه آن کدگذاری عصبی³ [3] در ش.ع.م ض.م مهم می شود.

جذابیت ش.ع.م برای محققان توانایی مغز در شناسایی الگوهای پیچیده و غیرخطی حتی در وجود نویز در داده ها یا محیط (مثل زمانی که در یک مکان پر سر صدا مانند مترو دو نفر که با هم در حال صحبت هستند ممکن است کلماتی از صحبت یک دیگر را به علت نویز محیطی متوجه نشوند ولی نهایتاً منظور

¹ Inhibitory

² excitatory

³ Neural Coding

یک دیگر را به طور کامل در می یابند) است. علت این قدرت شناسایی را می توان نتیجه چندین سطح پردازش شبکه ی ضربه مغز دانست.

از مهم ترین برتری های ش.ع.ض نسب به ش.ع.م امکان پیاده سازی سخت افزاری این شبکه هاست. همانطور که اشاره شده ضربه ها در زمان خلوت هستند پس انرژی کمتری برای انتشار اطلاعات مصرف می کنند[4]. اما در سمت مقابل ش.ع.م توان محاسباتی بالایی و طبعاً پردازنده های قدرتمندی نیاز دارند. پس با ش.ع.ض امکان ساخت سخت افزار کم توان و قادر به انجام کارهای پیچیده ممکن خواهد شد.

توالی ضربه ها معمولاً به صورت جمعی از توابع ضربه به همراه انرژی آن ضربه بیان می شوند و مشتق ناپذیر هستند. این موضوع عملاً استفاده از روش های آموزش مبتی بر مشتق را غیر ممکن می سازد. اگرچه روش هایی برای تخمین مشتق این توابع ضربه ارائه شده است[6] اما این تخمین ها پیچیدگی زیادی به همراه داشته و فرایند آموزش را سخت و بعضاً غیرممکن می کند.

یک سوال اصلی که پیش می آید این است که اگر روش های یادگیری مبتی بر مشتق قابل استفاده نیست پس نورون ها در ش.ع.ض چگونه قرار است آموزش ببینند؟ این سوالی است که در طی این پژوهش قرار است به آن پاسخ دهیم. اما در یک جمله کوتاه باید گفت با اینکه ادعا شده است که ش.ع.ض قدرتی برابر با ماشین تورینگ دارند[5] ولی آموزش آن ها هنوز یکی از مباحث باز و چالشی این حوزه است.

در ادامه این گزارش، در قسمت ۲ مروری به تحقیقات گذشته برای دستیابی به مدل های عمیق ش.ع.ض و توضیح ساختار و آموزش ش.ع.ض خواهیم داشت. سپس در قسمت ۳ به مطالعه ساختارهای عمیق جلورو¹ و کاملاً متصل² ش.ع.ض خواهیم پرداخت. و در قسمت ۴ شبکه های پیچشی ضربه ای³ را مشاهده خواهیم کرد در قسمت ۵ شبکه های بازگشتی ضربه ای⁴ را بررسی خواهیم کرد نهایتاً در قسمت ۶ با مطالعه استفاده از ش.ع.ض در یادگیری برخط این پژوهش را به پایان خواهیم رساند.

¹ Feedforward

² Fully conected

³ Spiking Convolutional Neural Network (SCNN)

⁴ Spiking recurrent Nerural Network (SRNN)

۲ - شبکه‌های عصبی ضربه‌ای _ ساختار و آموزش

ش.ع.ض نسل سوم شبکه‌های عصبی محسوب می‌شوند که باعث شد تعداد کثیری از محققان به روش‌های مبتنی بر کارکرد مغز در زمینه‌های مختلف مثل شناسایی الگو [7] روی آورند. ش.ع.ض در ابتدا با توجه به پتانسیل‌های گسسته الکتریکی گذرنده از سیناپس¹ پدید آمدند. در شبکه‌های زیستی، یک نورون وقتی شلیک می‌کند که انرژی غشایی نورون با توجه به مجموع انرژی‌های بازدارنده و وادارنده وارد شده بر سطح غشا از یک حد استانه‌ای مشخص بیشتر شود. نرخ، الگو و اطلاعات این شلیک شامل اطلاعات مربوط به محرک خارجی است. با این مقدمه آماده دقیق‌تر شدن در ساختار و روش یادگیری ش.ع.ض هستیم.

۱-۲ ساختار شبکه‌های ضربه‌ای

به طور عمومی ساختار یک ش.ع.ض از نورون‌های ضربه‌ای و اتصالات سیناپسی با وزن‌های تغییر پذیر تشکیل شده است. اولین گام برای پیاده‌سازی یک ش.ع.ض کدگذاری ورودی آنالوگ به یک توالی ضربه با روش‌های مبتنی بر نرخ [8]، کدگذاری زمانی [9] یا کدگذاری مبتنی بر جمعیت [10] است.

یک نورون زیستی در مغز، از سایر نورون‌ها توسط اتصالات سیناپسی ورودی می‌گیرد پس هم تغییرات عمل و هم تغییرات شبکه‌ای دارد. در ش.ع.ض تغییرات شبکه نسب به نورون زیستی بسیار ساده‌تر فرض شده است، بطوریکه نورون‌های مدل شده فقط از آستانه معمولی در مقابل آستانه‌های هیستریزیس و دیرگذاری که در مغز وجود دارد استفاده می‌شود. یک نورون پیش‌سیناپسی² سطح انرژی یک نورون پس‌سیناپسی³ را تغییر می‌دهد، به طوری که اگر به یک سطح آستانه مشخص برسید. باعث شلیک نورون پس‌سیناپسی می‌شود برای اولین بار هاجکین و هاکسلی موفق به مدل سازی این پدیده شدند [11]. در اصل آن‌ها مدلی از تولید ضربه مبتنی بر خواص مجرای یونی وابسته به ولتاژ ارائه کردند، که این مدل بر پایه سلول‌های عصبی یک نوع ماهی است.

¹ Synapse

² Presynaptic Neuron

³ Postsynaptic Neuron

بعد از هاجکین و هاکسلی مدل‌های دیگری مانند مدل واکنش ضربه¹[12] و مدل انتگرال سپس شلیک² ارائه شد که مدل دوم توجه بسیاری از محققان را به خود جلب کرده است.

گفتیم که نورون‌ها قرار است روی یک دیگر اثر وادارنده یا بازدارنده داشته باشند. قدرت این اثر در ارتباط سیناپسی توسط همان وزن رابطه سیناپسی بین نورون‌ها تعیین می‌شود. این وزن تغییر پذیر است و در طی فرایند آموزش تغییر خواهد کرد. قانون یادگیری ش.ع.ض مهم‌ترین و چالش‌برانگیزترین مورد برای طراحی این شبکه‌ها است.

۲-۲ قوانین یادگیری در شبکه‌های عصبی ضربه‌ای

منظور از یادگیری در اکثر شبکه‌های عصبی (چه ضربه‌ای چه غیر ضربه‌ای) تغییر وزن‌های تغییر پذیر در طول آموزش است. عمل شلیک نورون و تولید ضربه یادگیری خاصی دارد که در شبکه‌های غیرضربه‌ای غیرقابل استفاده است.

محققان علم اعصاب، انواع متنوعی از این نوع یادگیری را کشف کرده‌اند که همگی در زمره انعطاف بر پایه زمان ضربه³ یا به اختصار انعطاف ضربه قرار می‌گیرند. ویژگی اصلی این قانون یادگیری، اتصال دو نورون پیش و پس سیناپسی با اثر سیناپس (وزن) است که با توجه به زمان نسبی ضربه‌ها در دوره‌های چند ده میلی‌ثانیه‌ای تغییر می‌کند[13]. اطلاعات استفاده شده برای تغییر وزن هم از لحاظ زمانی و هم از لحاظ سیناپسی محلی است. در ادامه آموزش‌های با نظارت و بی‌نظارت در ش.ع.ض را بررسی می‌کنیم.

۱-۲-۲ یادگیری با نظارت در شبکه‌های عصبی ضربه‌ای

وقتی صحبت از یادگیری با نظارت می‌شود اولین نکته که در ذهن می‌آید وجود برچسب برای داده‌ها است. این برچسب‌ها برای محاسبه کردن خطا در شبکه مفید هستند. در شبکه‌های غیرضربه‌ای با توجه

¹ Spike Response Model

² Leaky integrate and fire

³ Spike timing dependent plasticity

با استفاده از گرادینان نزولی به دست آمده از خطای استفاده شده در یک تابع هزینه وزن‌ها تغییر می‌کنند. اما در ش.ع.ض یادگیری با نظارت تلاش در کمینه سازی خطا بین خروجی مورد نظر و توالی ضربه‌ها دارد که به خطای خواندن¹ در پاسخ به ورودی نیز معروف است.

از نقطه نظر زیستی، قابلیت پیاده‌سازی روش پس انتشارخطا در مغز همیشه مورد تردید و شک قرار داشته است. با تمرکز بر ش.ع.ض دو ایراد اساسی در فرمول زیر که فرمول اصلی روش پس انتشارخطاس و از قانون زنجیره‌ای به دست آمده و در تمامی روش‌های مبتنی بر پس انتشارخطا وجود دارد (بیشاپ، ۱۹۹۵)، دیده می‌شود.

$$\delta_j^u = g'(a_j^\mu) \sum_k \omega_{kj} \delta_k^\mu \quad (1)$$

که در بالا δ_k^μ و δ_j^μ مشتق‌های جزئی تابع هزینه برای ورودی μ برای دو نورون دلخواه j یا k هستند. نورون j به عنوان نورن مقصد در نظر گرفته می‌شود و مجموعه‌ای از نورون‌ها که با k شاخص دهی می‌شوند به عنوان ورودی آن در نظر گرفته می‌شود. تابع g نیز به عنوان تابع فعالیت شبکه در نظر گرفته می‌شود.

دو قسمت از رابطه شماره (۱) برای ش.ع.ض مشکل ساز است. مشکل اول، مشتق تابع g وجود ندارد، چون که ورودی تابع جمع توابع دلتای دیراک است. مشکل دوم که هم در شبکه‌های ضربه‌ای و هم در شبکه‌های غیرضربه‌ای وجود دارد مشکل جابجایی وزن است [14] به طور خلاصه اینکه فرض کرده‌ایم که وزن‌های یکسان و متقارن در مسیر فیدبک موجود هستند تا معادله (۱) قابل استفاده شود.

مشکل اول معمولاً با تخمین مشتق حل می‌شود (البته واضح است که این کار مبنای زیستی ندارد). در مورد مشکل دوم، پیشرفت‌هایی در این زمینه صورت گرفته است، مثلاً اینکه از وزن‌های تصادفی فیدبک استفاده کنیم.

¹ Readout error

از مدل‌هایی که اخیراً به طور خاص برای یادگیری بانظارت معرفی شده است و عملکرد قابل توجه‌ای دارد ماشین فیمن¹ است [15]. در این مدل فرض شده است که ش.ع.ض تمایل دارد که از اصل بقای بار² پیروی کند این تمایل باعث می‌شود که توزیع اتصالات نه تنها بسته به محرک‌های ورودی به نورون‌های ورودی، بلکه به محرک‌های خروجی از نورون‌های خروجی نیز تعیین می‌شود که گویی راه‌حلی از روش ابتدایی محدود در یک سیستم سیال است.

۲-۲-۲ یادگیری بی نظارت در شبکه‌های عصبی ضربه‌ای

استفاده از مدل توضیح داده شده در بخش یادگیری با نظارت بنا به دلایلی مانده و اگرایی اتصالات امکان پذیر نیست در این بخش مروری به روش‌های به کار گرفته شده برای یادگیری بی نظارت ش.ع.ض که در آن برجستگی برای داده‌ها نداریم، خواهیم داشت.

در یک نگرش ساده انعطاف ضربه، اگر یک نورون پیش‌سیناپسی کمی قبل‌تر (حدود ده میلی‌ثانیه) از نورون پس‌سیناپسی شلیک کند، وزن اتصال آن دو قوی‌تر و منطقی خواهد شد. اگر نورون پیش‌سیناپسی کمی بعد از نورون پس‌سیناپسی شلیک کند، رابطه بین آن دو نورون منطقی و درست نیست و ضعیف‌تر خواهد شد. قوی شدن این اتصال قوی شدن طولانی مدت³ و ضعیف شدن ارتباط تنزل طولانی مدت⁴ نامیده می‌شود. علت استفاده از عبارت طولانی مدت، تمایز قائل شدن بین حالت گذرا در مقایس میلی‌ثانیه است.

فرم ایده‌آل قانون انعطاف ضربه در فرمول (۲) آمده است، این فرم در آزمایش‌های مختلف برای یک جفت ضربه حاصل شده است [16].

$$\Delta w = \begin{cases} A e^{-\frac{|t_{pre}-t_{post}|}{\tau}} & t_{pre} - t_{post} \leq 0, A > 0 \\ B e^{-\frac{|t_{pre}-t_{post}|}{\tau}} & t_{pre} - t_{post} > 0, B < 0 \end{cases} \quad (2)$$

w در فرمول (۲) نشان دهنده وزن سیناپسی است، همچنین A و B به طور معمول ثابت و پارمترهای قابل یادگیری هستند. τ ثابت زمانی (معمولاً ۱۵ میلی‌ثانیه) پنجره یادگیری است. عبارت نخست بیان گر قوی

¹ Feynman machine model

² charge conservation principle

³ Long term potentiation

⁴ Long term depression

شدن و عبارت دوم بیان گر تنزل است. البته انواع دیگر و گسترش‌های این فرمول برای ارضا یه محدودیت ریاضی بیشتر در کاربرد مورد استفاده قرار می‌گیرد.

می‌توان یادگیری بی‌نظارت را به صورت احتمالاتی نیز نگریست. بسیاری از مطالعات نشان داده‌اند که حداقل یک تحلیل تقریبی بیزی از محرک‌های حسی در مغز رخ می‌دهد [17] [18] [19]. در استنتاج بیزی، علل پنهان (مانند حضور یک شی از یک دسته خاص) با استفاده از دانش قبلی و احتمال مشاهدات جدید برای به دست آوردن یک احتمال پسین از علت احتمالی استنباط می‌شوند. محققان نقش احتمالی محاسبات احتمالی (بیزی) را به عنوان یک مرحله پردازش اطلاعات اولیه در مغز از نظر انعطاف زمانی ضربه در نظر گرفته‌اند. نسلر یک فرم از انعطاف زمانی ضربه را نشان داد که با نوروهای ورودی ضربه پواسون همراه با مدار تصادفی برنده همه چیز¹ استفاده می‌کند، این فرم قادر است یک الگوریتم بیشینه سازی انتظار آنلاین² را برای یادگیری پارامترهای یک توزیع مخلوط چند جمله ای تقریب بزند. قانون انعطاف زمانی ضربه استفاده شده برای این شبکه‌ها در فرمول (۳) آمده است.

$$\Delta w_{ki} = \begin{cases} e^{-w_{ki}} - 1, & 0 < t_k^f - t_i^f < \epsilon \\ -1, & \text{otherwise} \end{cases} \quad (3)$$

همان طور که از فرمول پیداس در اینجا هم اگر تفاضل زمان شلیک نورو پس‌سناسپی از شلیک نورو پیس‌سناسپی بزرگ تر از صفر و کوچک تر از یک عددی (مثلا $\epsilon = 10\text{ms}$) بود با توجه به وزن سیناسپی قبلی وزن بروزرسانی می‌شود.

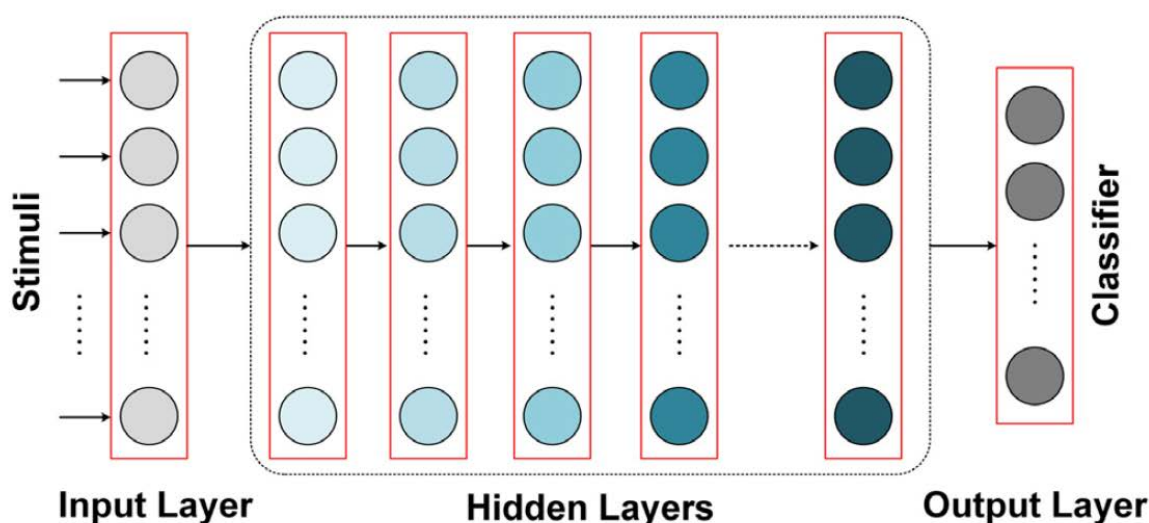
قانون مذکور مورد توجه بسیاری از محققان قرار گرفته است. از جمله مطالعاتی که با توجه به این قانون انجام گرفته می‌توان به کار رزنده و همکارانش که یک قانون یادگیری زیستی قابل قبول مبتنی بر توزیع مشترک ادراکات و دلایل پنهان برای تطبیق توالی‌های ضربه خود به خود برای مطابقت با توزیع تجربی توالی‌های ضربه واقعی ایجاد کردند [20]، اشاره کرد.

¹ Winner-take-all

² online expectation maximization

۳- یادگیری عمیق در شبکه‌های عصبی ضربه‌ای

یادگیری عمیق زمانی رخ می‌دهد که در ساختار شبکه از چندین لایه استفاده شود. این یادگیری عملکرد خارق‌العاده در حوزه یادگیری ماشین و هوش مصنوعی دارد [21]. ساختار عصبی چندلایه در مغز پستانداران الهام‌بخش محققان است تا تمرکز خود را روی عمیق شبکه و نورون‌های غیرخطی به جای شبکه‌های کم عمق ولی با عرض زیاد معطوف کنند. هم از نظر تئوری و هم در عمل شبکه‌های عمیق عملکرد بهتری نسبت به شبکه‌های پهن دارند [22]. شبکه‌های عمیق ویژگی‌های پیچیده‌تری برای الگوریتم پس‌انتشار خطا استخراج می‌کنند. شکل ۱، ساختار یک شبکه عصبی عمیق با چندین لایه مخفی را نشان می‌دهد.



شکل ۱، ساختار یک شبکه عمیق ساده، کاملاً متصل، با لایه‌های ورودی، مخفی و خروجی

لایه ورودی به مرور یاد می‌گیرد که پیش‌پردازش لازم را روی داده‌های ورودی انجام

دهد. لایه‌های مخفی قابلیت استخراج ویژگی‌های پیچیده‌تر را دارند و وظیفه لایه

آخر عمل دسته‌بندی است. [24]

در بیشتر مسائل دسته‌بندی، لایه خروجی شبکه معمولاً از سافت‌مکس¹ بهره می‌برد تا بردار خروجی تبدیل به یک بردار احتمال گردد که در اصل نشان دهنده محتمل بودن ورودی به عضویت در هر یک از کلاس‌های مسئله است.

¹ Softmax

ش.ع.ض نتایج بسیار امیدبخشی در وضایف بازشناسی الگو نشان دادند [23]. اما عملکرد ش.ع.ض عمیق که به صورت مستقیم آموزش دیده‌اند از شبکه‌های عمیق معمول پایین‌تر است. بنابراین آموزش شبکه‌های عمیق ضربه‌ای که دقتی در سطح شبکه‌های عمیق معمولی داشته باشند موضوعی چالش برانگیز است، چراکه پیاده‌سازی سخت‌افزاری در شبکه‌های عمیق بسیار مهم است.

۳-۱ شبکه‌های کاملاً متصل ضربه‌ای

در تحقیقات جدید، چندین شبکه عصبی ضربه‌ای عمیق با استفاده از انعطاف ضربه و گرادین نزولی تصادفی^۱ آموزش داده شدند. این شبکه به موفقیت‌های زیادی در زمینه شناسایی الگو [25] و دقت ۹۵٪ در دسته‌بندی دادگان ام‌نیست^۲ دست یافته‌اند. اوکانر و ویلنگ الگوریتمی برپایه پس‌انتشارخطا معرفی کردند [26] که با استفاده از ضرب خارجی ضربه‌ها، شبکه آموزش می‌بیند. این شبکه با استفاده از تابع فعالیت^۳ ReLU دقت ۹۸٪ را در دسته‌بندی دادگان ام‌نیست حاصل می‌کند. لی و همکارانش نیز پیشنهاد رفتار با پتانسیل غشائی به عنوان یک سیگنال مشتق‌پذیر را دادند [27] تا همانند یک نورون غیرخطی در شبکه‌های عصبی معمولی باشند، حاصل این تحقیق دقت ۹۸/۸۸٪ برای دسته‌بندی دادگان ام‌نیست بود، این درحالی است که حجم محاسباتی حدود یک پنجم شبکه غیرضربه‌ای با دقت مشابه است. شکل ۲ نمایی از ساختار این شبکه کار نشان می‌دهد.

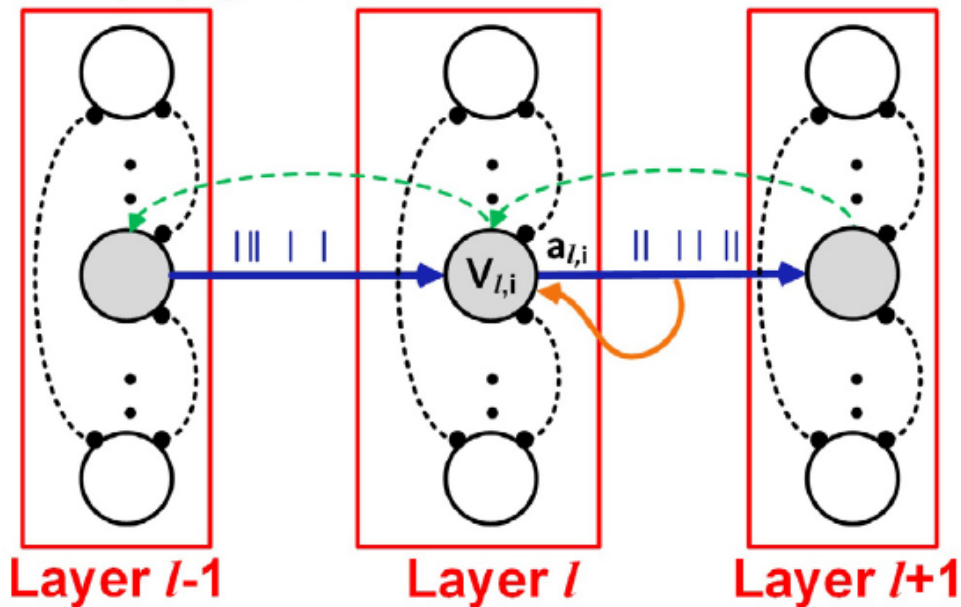
یک رویکرد سخت‌افزار محور برای داشتن ش.ع.ض قدرتمند استفاده از مدل ازپیش‌آموزش‌دیده شده‌ی غیرضربه‌ای در آموزش شبکه‌ی ضربه‌ای است (تبدیل ش.ع.م به ش.ع.ض)، این امر در پیاده‌سازی سخت‌افزاری موثر است [28]. برای جایگزینی مقادیر فعال‌سازی ممیز شناور در ش.ع.م، معمولاً از کدگذاری مبتنی بر نرخ استفاده می‌شود که در آن فعال‌سازی‌های بالاتر با نرخ‌های افزایشی بالاتر جایگزین می‌شوند. با استفاده از این رویکرد، چندین مدل توسعه یافته است که عملکرد دقت بسیار خوبی را به دست آورده است [29][30][31].

^۱ stochastic gradient descent (SGD)

^۲ Mnist

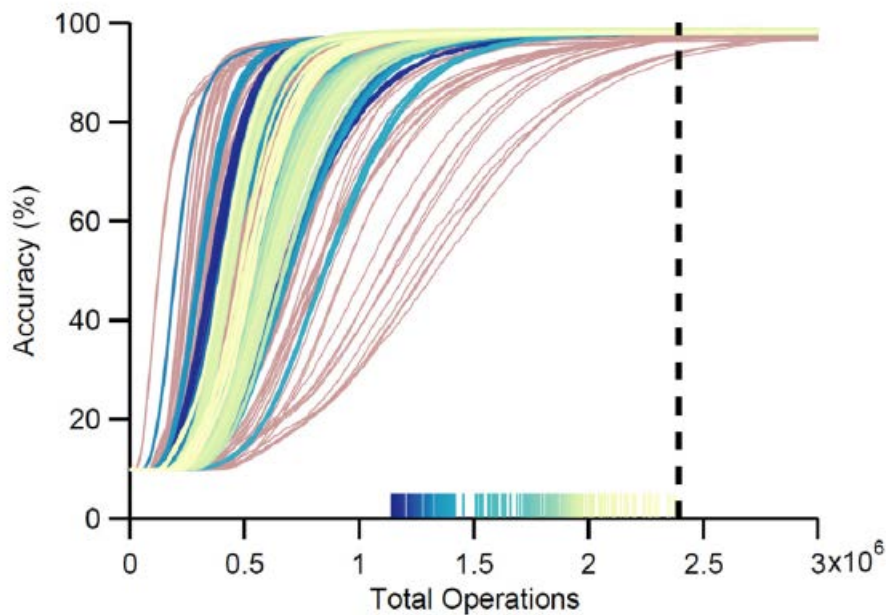
^۳ Rectified linear unit

Reset Signal
Spike Train
Lateral Inhibition
Back-propagation



شکل ۲ مقدار فعال سازی نرون $a_{l,i}$ ، توسط پتانسیل غشایی نرون داده می شود. تابع فعال سازی متمایز، که توسط ورودی تحریکی نرون، مهار جانبی و آستانه محاسبه می شود، برای توسعه پس انتشار با استفاده از قانون زنجیره استفاده می شود. مقدار فعال سازی خروجی لایه جاری (لایه l) به عنوان ورودی لایه بعدی در الگوریتم پس انتشار استفاده می شود [24].

در مطالعه ی دیگر برای برآورد سخت افزاری ش.ع.ض عمیق [32] که از ساختار شبکه مشابه شکل ۲ استفاده می کند ولی تمرکز بر استفاده از پارمترهای بهینه تر و تنظیم کردن ش.ع.ض که در نتیجه کاهش مصرف و تاخیر مدل در تبدیل ش.ع.م. به ش.ع.ض حاصل شده است. در شکل ۳ دقت حاصل در آزمایش های مختلف برای مقدار پارمترها نشان داده شده است.



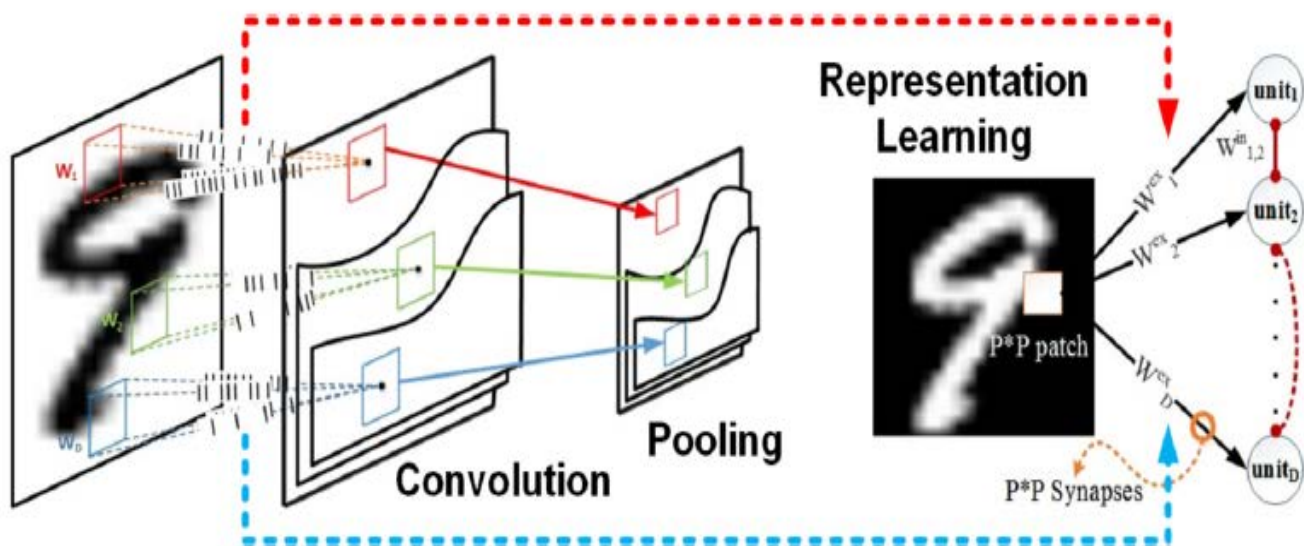
شکل ۳، آزمایش‌های مختلف برای مقدار پارمترها و دقت حاصل شده برای شبکه کاملاً متصل

۴- شبکه‌های عصبی پیچشی ضربه‌ای

کاربرد شبکه‌های عصبی عمیق پیچشی بیشتر در زمینه‌های پردازش تصویر است. ساختار این شبکه از چندین سری کانولوشن و رای‌گیری^۱ تشکیل شده است. فیلترهای مدل شده در شبکه‌های پیچشی را می‌توان با قانون یادگیری انعطاف ضربه ترکیب کرد تا شبکه‌های عصبی پیچشی ضربه‌ای تولید شود.

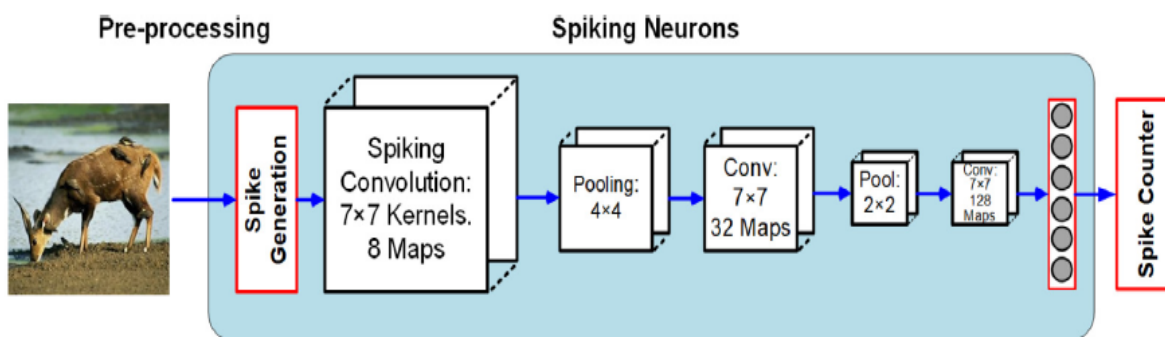
شبکه‌های پیچشی معمولاً از یک میدان دریافت در لایه‌های اولیه برای استخراج ویژگی‌ها از محرک‌ها با پیچاندن میدان دریافت بر روی ورودی (مثلاً تصویر) استفاده می‌کنند. لایه‌های بعدی فضای نتیجه شد از لایه قبلی را برای یادگیری ویژگی‌های محرک پیچیده و انتزاعی ترکیب می‌کنند. بازنمایی‌های (آموزش داده شده یا مهندسی شده) و قوانین یادگیری انعطاف ضربه را می‌توان برای توسعه شبکه‌های پیچشی ضربه‌ای استفاده کرد. شکل ۴ یک لایه کانولوشنی و رای‌گیری آموزش داده شده با ضربه‌های محلی را نشان می‌دهد.

¹ Pooling



شکل ۴، وزن های سیناپسی تحریکی متصل به نورون ها در لایه نمایش، فیلترهای کانولوشنی را مشخص می کنند. این معماری تعیین می کند که یادگیری بازنمایی در ش.ع.ض تک لایه می تواند برای آموزش شبکه های پیچشی ضربه ای چند لایه ای مورد استفاده قرار گیرد [24].

همانطور که در بخش ۳-۱ گفته شد، یکی از رویکردهای محبوب استفاده از یک ش.ع.م. از پیش آموزش دیده است برای استفاده در ش.ع.ض است. برای ش.ع.ض پیچشی نیز می توانیم از یک شبکه ی پیچشی آموزش دیده شده است کنیم. مطالعات زیادی نشان داده اند که استفاده از یک ش.ع.ض پیچشی تبدیل شده عملکردی نزدیک به یک شبکه ی پیچشی معمولی دارد که البته حجم محاسباتی خیلی کمتری را می طلبد. ساختار یکی از اولین شبکه های تبدیلی پیچشی به ضربه ای پیچشی را در شکل ۵ مشاهده می کنید.



شکل ۵، تصویر ورودی، پس از پیش پردازش، بر اساس شدت پیکسل به توالی‌های ضربه تبدیل می‌شود. لایه‌های ضربه‌ای از وزنه‌هایی استفاده می‌کنند که توسط یک شبکه‌ی پیچشی غیرضربه‌ای آموزش داده شده است. آخرین عنصر شبکه، نرونی که حداکثر فعالیت (فرکانس ضربه) را دارد به عنوان کلاس تصویر انتخاب می‌کند.

۵ - شبکه‌های عصبی بازگشتی ضربه‌ای

به زبان ساده یک شبکه عصبی را می‌توان بازگشتی تلقی کرد اگر در گراف ساختار آن حداقل یک دور دیده شود. خاصیت این دور ایجاد حافظه در شبکه است. شبکه‌های عصبی بازگشتی برای تحلیل داده‌های زمانی مورد استفاده قرار می‌گیرند. این شبکه‌ها اکثراً با استفاده از پس انتشارخطا در زمان آموزش می‌بینند، که مدل را برای چندین قدم در زمان باز کرده و به آموزش مدل به روش معمولی می‌پردازد. این شبکه‌ها در صورت عمیق شدن دچار مشکل محو شدن یا انفجار گرادیان می‌شوند [33]. برای جلوگیری از مشکل بیان شده باید از ضرب شدن پی در پی مشتق‌ها جلوگیری شود. شبکه‌های حافظه طولانی کوتاه مدت¹ [34] از شبکه‌های بازگشتی محبوبی هستند که از نرونها ضربه‌ای استفاده نمی‌کنند.

فقط تعداد انگشت شماری از شبکه‌های بازگشتی ضربه‌ای وجود دارد که آن‌ها هم از رویکرد محبوب تبدیل شبکه‌های بازگشتی غیرضربه‌ای به شبکه‌های بازگشتی ضربه‌ای بدست آمده‌اند. شبکه‌ای که از تبدیل شبکه المن حاصل شد جز اولین شبکه‌های طراحی شده است [4].

¹ Long short-term memory (LSTM)

۶ - یادگیری بر خط در شبکه‌های عصبی ضربه‌ای

در ابتدا باید به تفاوت یادگیری برخط با یادگیری معمولی توجه کنیم. در یادگیری معمولی ما همه دادگان را در ابتدا داریم و توضیح این دادگان ثابت است. پس می‌توان با کل یا بخشی از دادگان به آموزش شبکه پرداخت اما در یادگیری برخط ما فقط یک یا بخش کوچکی از دادگان را در ابتدا در اختیار داریم و به مرور زمان داده‌های جدید که ممکن است توزیع متفاوتی نیز داشته باشند وارد می‌شوند حجم و سرعت وارد شدن داده‌های جدید نیز در مضایف مختلف می‌تواند متفاوت باشد. استفاد از ش.ع.م و روش‌های سنتی تر نیاز با الگوریتم‌های برای تطبیق مدل با توزیع داده‌ای جدید دارد که البته این الگوریتم‌ها هزینه‌بر هستند و استفاده از آن‌ها در وظایف دادگان حجیم عملاً ممکن نیست.

از ویژگی‌های بسیار مهم ش.ع.ض عملکرد مناسب آن‌ها در یادگیری برخط است. جریان‌های داده ممکن است وابستگی‌های زمانی بین برجسب‌های کلاس نشان دهند، یعنی اگر برجسب نمونه فعلی y_t باشد ممکن است تحت تأثیر برجسب‌های نمونه قبلی $(y_{t-1}, y_{t-2}, \dots)$ قرار گرفته باشد. وابستگی زمانی می‌تواند به تعیین چگونگی ارتباط ویژگی‌های ورودی با یکدیگر در طول زمان کمک کند. ش.ع.ض از نمایش اطلاعات اسپایک استفاده می‌کنند تا قوانین یادگیری در زمان افزایش را بسازند که توانایی ثبت ارتباط زمانی بین متغیرهای زمانی در جریان داده را نشان می‌دهد.

استفاده از برخی ش.ع.ض در یادگیری برخط امکان افزایش سرعت و کاهش پیچیدگی محاسباتی فرآیند یادگیری را فراهم می‌کند، با توجه به محلی بودن یادگیری پیاده سازی موازی آن مناسب است. در جدول ۱ برخی از ش.ع.ض که برای یادگیری برخط توسعه داده شده‌اند آمده است.

نام شبکه	نقاط قوت	نقاط ضعف
SpikeProp	(۱) قادر به حل مسائل پیچیده طبقه بندی است (۲) از نظر محاسباتی قدرتمند است	(۱) هنگامی که نورون پس سیناپسی دیگر برای هر الگوی ورودی شلیک نمی‌کند، مکانیزمی برای "تقویت" وزن های سیناپسی وجود ندارد. (۲) فقط اولین ضربه تولید شده توسط یک نورون مرتبط است و بقیه دوره زمانی نورون نادیده گرفته می شود. (۳) برای استفاده در یادگیری برخط بسیار کند است

		<p>(۴) تعداد زیاد اتصالات سیناپسی، زمانی که مجموعه داده با ابعاد بالا در نظر گرفته می شود.</p> <p>(۵) آنها اغلب از نظر زیستی غیر قابل قبول در نظر گرفته می شوند زیرا به گسترش غیر محلی سیگنال های خطا از یک سیناپس به سیناپس دیگر نیاز دارند.</p>
SHL	<p>(۱) احتمالاً سراسر ترین راه حل را از دیدگاه زیستی واقع بینانه ارائه می دهد</p>	<p>(۱) در طول آموزش جریان های سیگنال همه شلیک های ناخواسته را سرکوب می کنند، تنها همبستگی فعالیت های پیش سیناپسی و پس سیناپسی در حول و حوش زمان شلیک هدف اتفاق می افتد. در موارد دیگر، این همبستگی وجود ندارد و هیچ مکانیزی برای تضعیف وزن های سیناپسی که نورون را در زمان های نامطلوب در مرحله آزمایش شلیک می کند، وجود ندارد.</p> <p>(۲) سیناپس ها وزن خود را تغییر می دهند حتی زمانی که نورون دقیقاً در زمان های مورد نظر شلیک می کند. بنابراین، SHL تنها با افزودن محدودیت های اضافی یا قوانین یادگیری بیشتر می تواند به راه حل های پایدار دست یابد</p>
ReSuMe	<p>(۱) مشکلات SpikeProp و قانون SHL را حل می کند</p> <p>(۲) مستقل از مدل های نورونی مورد استفاده است</p> <p>(۳) می تواند به طور موثر الگوهای زمانی و مکانی-زمانی دلخواه را بیاموزد</p> <p>(۴) فرآیند یادگیری به سرعت همگرا می شود</p>	<p>(۱) با وجود ادعای مناسب بودن برای یادگیری برخط، ساختار شبکه مورد استفاده ثابت است و با محرک های ورودی سازگار نمی شود.</p> <p>(۲) قادر به پیش بینی ورودی ها پس از یک بار ارائه نمونه های آموزشی نیست</p> <p>(۳) اگرچه از انعطاف ضربه استفاده می کند که از نظر بیولوژیکی قابل قبول است، ولی طبیعت محلی آن ظرفیت یادگیری آن را محدود می کند</p>
Spiketemp	<p>(۱) قادر است ورودی ها را تنها پس از یک بار ارائه نمونه های آموزشی پیش بینی کند</p> <p>(۲) برای طیف گسترده ای از مجموعه داده ها با ابعاد و تعداد کلاس های مختلف مقیاس پذیر است</p>	<p>(۱) هزینه زمانی در مقایسه با رویکرد مبتنی بر رتبه افزایش می یابد</p>

	(۳) کارآمدتر از روش یادگیری مرتبه‌ای موجود است	
Tempotron	(۱) می‌تواند دسته‌بندی طیف وسیعی از کلاس‌های ورودی را بیاموزد که در آن اطلاعات دسته‌بندی در تعداد ضربه‌ها رمزگذاری نمی‌شوند، بلکه در زمان‌های تأخیر ضربه‌های منفرد یا در الگوهای همگام‌سازی به صورت جفتی و مرتبه بالاتر کدگذاری می‌شوند.	(۱) فقط می‌توان آن را برای شبکه‌های تک لایه اعمال کرد (۲) به خروجی 0 یا 1 ضربه در یک بازه زمانی از پیش تعیین شده محدود می‌شود، و بنابراین خروجی اطلاعات زمان بندی اسپیک را رمزگذاری نمی‌کند.
eSNN	(۱) مدل عصبی آن امکان شبیه سازی بلادرنگ بسیار سریع شبکه‌های بزرگ و هزینه محاسباتی کم را فراهم می‌کند. (۲) ماهیت در حال تکامل آنها (نرون‌های پراکنده به طور تدریجی در طول زمان برای استنباط الگوهای زمانی از داده‌ها تکامل می‌یابند) امکان انباشت دانش را با رسیدن داده‌ها بدون ذخیره و آموزش مجدد مدل با داده‌های گذشته فراهم می‌کند. (۳) آنها برای محیط‌های غیر ثابت توصیه می‌شوند، زیرا تغییرات در داده‌های جریان ورودی بلافاصله به عنوان رویدادهای باینری یا ضربه کدگذاری می‌شوند، که یکی از مناسب ترین استراتژی‌های رمزگذاری داده برای تطبیق با رانش‌ها است.	(1) تعداد نورون آن به طور نامحدود با هر نمونه جدید در یادگیری برخط رشد می‌کند. (2) نورون‌ها را با میانگین وزن‌های تخمین زده شده با استفاده از ترتیب رتبه به روز می‌کند/ادغام می‌کند. این امکان وجود دارد که دو الگوی ورودی دارای رتبه یکسانی باشند (ضربه‌ها به ترتیب یکسان رخ دهند) اما زمان‌های دقیق شلیک آن‌ها از هم فاصله زیادی دارد. این ممکن است منجر به ش.ع.ض با تعداد نورون‌های بیشتر و از دست دادن دانش ذخیره شده در شبکه شود.

جدول ۱، بررسی مدل‌های ضربه‌ای توسعه داده شده برای یادگیری برخط

۷ - نتیجه گیری

شبکه‌های عصبی ضربه‌ای به عنوان نسل سوم شبکه‌های عصبی تلاشی برای مدل سازی آنچه که واقعا در مغز انسان اتفاق می‌افتد است. با معرفی این شبکه‌ها علاقه محققان برای استفاده از آن‌ها در انواع یادگیری، انواع شبکه‌ها (پیچشی، بازگشتی و...) و انواع وظایف (پردازش گفتار، پردازش متن، پردازش تصویر و...) مهم ترین چالش در این شبکه‌ها قانون یادگیری است، چون تابع فعالیت در این شبکه‌ها مشتق پذیر نیست و استفاده از روش‌های مبتنی بر گرادیان به طور مستقیم ممکن نیست، هرچند روش‌های مانند تخمین مشتق و انعطاف زمانی ضربه تلاشی برای یادگیری در این شبکه‌ها است، ولی پیدا کردن یک قانون یادگیری مناسب به صورت عمومی هنوز یک چالش باز است. مهم‌ترین انگیزه محققان برای روی آوردن به این شبکه‌ها امکان پیاده سازی آن‌ها با سخت افزارهای کم‌توان است.

لیست مراجع

- [1] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern*, 2016.
- [2] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. -r. Mohamed, N. Jaitly and e. al., "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," *IEEE Signal Processing Magazine*, pp. 82-97, 2012.
- [3] W. Bair and C. Koch, "Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey," *Neural Computation*, pp. 1185-1202, 1996.
- [4] J. V. Stone, Principles of neural information theory: Computational neuroscience and metabolic efficiency, Sebtel Press, 2018.

- [5] W. Maass, "Lower bounds for the computational power of networks of spiking neurons," *Neural Computation*, pp. 1-40, 1996.
- [6] D. Huh and T. J. Sejnowski, "Gradient descent for spiking neural networks," 2017. [Online]. Available: <http://arxiv.org/abs/1706.04698>.
- [7] S. Ghosh-Dastidar and H. Adeli, "Spiking neural networks," *International Journal of Neural Systems*, pp. 295-308, 2009.
- [8] W. Gerstner and W. M. Kistler, *Spiking neuron models: Single neurons, populations, plasticity*, Cambridge University Press, 2002.
- [9] S. M. Bohte, "The evidence for neural information processing with precise spike-times: A survey," *Natural Computing*, pp. 195-206, 2004.
- [10] S. M. Bohte, H. La Poutre and J. N. Kok, "Unsupervised clustering with spiking neurons by sparse temporal coding and multilayer RBF networks," *IEEE Transactions on Neural Networks*, pp. 426-435, 2002b.
- [11] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *The Journal of Physiology*, pp. 500-544, 1952.
- [12] R. Jolivet, J. Timothy and W. Gerstner, "The spike response model: A framework to predict neuronal spike trains," in *In Artificial neural networks and neural information processing—ICANN/ICONIP*, 2003.
- [13] N. Caporale and Y. Dan, "Spike timing-dependent plasticity: A Hebbian learning rule," *Annual Review of Neuroscience*, pp. 25-46, 2008.

- [14] S. Grossberg, "Competitive learning: From interactive activation to adaptive resonance," *Cognitive Science*, pp. 23-63, 1987.
- [15] Cho, M.W. Supervised learning in a spiking neural network. *J. Korean Phys. Soc.* 79, 328–335 (2021). <https://doi.org/10.1007/s40042-021-00254-4>
- [16] Y. Dan and M. -M. Poo, "Spike timing-dependent plasticity: From synapse to perception," *Physiological Reviews*, pp. 1033-1048, 2006.
- [17] Doya, K. (2007). Bayesian brain: Probabilistic approaches to neural coding. MIT press. Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., et al. (2012). A large-scale model of the functioning brain. *Science*, 338(6111), 1202–1205.
- [18] Kröger, B. J., Kannampuzha, J., & Neuschaefer -Rube, C. (2009). Towards a neurocomputational model of speech production and perception. *Speech Communication*, 51(9), 793–809.
- [19] Mozer, M. C., Pashler, H., & Homaei, H. (2008). Optimal predictions in everyday cognition: The wisdom of individuals or crowds? *Cognitive Science*, 32(7), 1133–1147
- [20] Brea, J., Senn, W., & Pfister, J. -P. (2011). Sequence learning with hidden units in spiking neural networks. In *Advances in neural information processing systems* (pp. 1422–1430).
- [21] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," *Nature*, pp. 436-444, 2015a.
- [22] I. Goodfellow, Y. Bengio and A. Courville, *Deep learning*, MIT Press, 2016.
- [23] S. Ghosh-Dastidar and H. Adeli, "Spiking neural networks," *International Journal of Neural Systems*, pp. 295-308, 2009.

- [24] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier and A. Maida, "Deep learning in spiking neural networks", *Neural Netw.*, vol. 111, pp. 47-63, Mar. 2019.
- [25] J. M. Brader, W. Senn and S. Fusi, "Learning real-world stimuli in a neural network with spike-driven synaptic dynamics," *Neural Computation*, p. 2881–2912, 2007.
- [26] P. O'Connor and M. Welling, "Deep spiking networks," 2016. [Online]. Available: <https://arxiv.org/abs/1602.08323>.
- [27] J. H. Lee, T. Delbruck and M. Pfeiffer, "Training deep spiking neural networks using backpropagation," *Frontiers in Neuroscience*, p. 508, 2016.
- [28] Perez-Carrasco, J., Zhao, B., Serrano, C., Acha, B., Serrano-Gotarredona, T., Chen, S., et al. (2013). Mapping from frame-driven to frame-free event-driven vision systems by low-rate rate-coding and coincidence processing. application to feed forward convnets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35, 2706–2719
- [29] Diehl, P. U., Neil, D., Binas, J., Cook, M., Liu, S. -C., & Pfeiffer, M. (2015). Fastclassifying, high-accuracy spiking deep networks through weight and threshold balancing. In *Neural networks, 2015 international joint conference on* (pp. 1–8). IEEE
- [30] Esser, S. K., Appuswamy, R., Merolla, P., Arthur, J. V., & Modha, D. S. (2015). Backpropagation for energy-efficient neuromorphic computing. In *Advances in neural information processing systems* (pp. 1117–1125)
- [31] Rueckauer, B., Hu, Y., Lungu, I. -A., Pfeiffer, M., & Liu, S. -C. (2017). Conversion of continuous-valued deep networks to efficient event-driven networks for image classification. *Frontiers in Neuroscience*, 11, 682.
- [32] Neil, D., & Liu, S. -C. (2016). Effective sensor fusion with event-based sensors and deep network architectures. In *Circuits and systems, 2016 IEEE international symposium on* (pp. 2282–2285). IEEE.

- [33] Y. Bengio, P. Simard and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, p. 157–166, 1994.
- [34] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, p. 1735–1780, 1997.
- [35] P. U. Diehl and M. Cook, "Unsupervised learning of digit recognition using spike-timing-dependent plasticity," *Frontiers in Computational Neuroscience*, pp. 1-9, 2015.