نام و نام خانوادگی: درسا الهوئی

شماره دانشجویی: ۹۷۲۲۲۰۰۵

عنوان : بخش چهارم تمرین سوم علوم اعصاب محاسباتی

۱. مدل R-STDP

R-STDP مزایای یادگیری تقویت شده و STDP را با یکدیگر ترکیب می کند.

نورون ها در مغز توسط سیناپس ها به یکدیگر متصل شده اند که قدرت این اتصال می تواند با گذشت زمان، ضعیف یا قوی شود. STDP نوعی یادگیری بدون نظارت است که در مناطق مختلف مغز به ویژه در قشر بینایی مشاهده می شود. STDP با در نظر گرفتن اختلاف زمانی بین اسپایک های pre/post synaptic عمل می کند. STDP در پیدا کردن ویژگی های تکرار شونده از نظر آماری به خوبی عمل می کند با این حال، به عنوان یک الگوریتم یادگیری بدون نظارت در تشخیص ویژگی های نادر اما قابل تشخیص دچار سختی می شود.

R-STDP که یک قانون یادگیری است شامل یک سیگنال جایزه در ترکیب با STDP می باشد، اخیرا مورد تمرکز تحقیقات واقع شده است. این روش قصد دارد تا عملکرد تنظیم کننده های عصبی مانند دوپامین که در مغز ساطع می شوند را تقلید کند. بنابراین RSTDP میتواند برای robot control بسیار مفید باشد زیرا ممکن نیاز به سیگنال آموزشی خارجی را ساده می کند و منجر به کارهای پیچیده تری می شود.

برای feed کردن دیتا به SNN، دیتا باید به نوعی تبدیل به اسپایک شود. به علاوه جایزه باید با دقت به SNN اختصاص یابد و اگر مقدار آن خیلی کم و یا خیلی زیاد باشد یادگیری را بی ثبات می کند. وزن های شبکه نیز عامل مهمی برای یادگیری می باشند چراکه اگر وزن ها مناسب نباشند، ممکن است پروسه یادگیری زمان زیادی صرف کند و یا در کل دچار شکست شود.

در دیتاست های مختلف تصاویر نشان داده شد که R-STDP ویژگی های بصری متمایز را استخراج می کند در حالی مدل بدون نظارت و کلاسیک STDP هر ویژگی ای که مداوم تکرار می شود را استخراج می کند. در نتیجه در این چنین دیتاست ها، R-STDP عملکرد بهتری نسبت به STDP داشت. علاوه بر این R-STDP برای یادگیری آنلاین مناسب است و می تواند با تغییرات شدید سازگار شود. لازم به ذکر است که هر دو عمل استرج ویژگی و کلاس بندی توسط اسپایک ها انجام می شوند (هر نورون یک اسپایک). بنابراین شبکه از نظر سخت افزاری سازگار و دارای انرژی کم است.

چندین مطالعه نشان داده است سیستم پاداشی مغز نقش مهمی تصمیم گیری و شکل گیری رفتارها دارد. این همچنین به عنوان یادگیری تقویتی (RL) نیز شناخته می شودکه به وسیله آن یادگیرنده تشویق می شو تا رفتارهای پاداش آور را تکرار و از رفتارهایی که منجر به مجازت می شود، دوری کند. مشخص شده است که دوپامین به عنوان یک تنظیم کننده عصبی، یکی از مواد شیمیایی مهم در سیستم پاداش است که ترشح آن متناسب با پاداش مورد انتظار در آینده است. همچنین نشان داده شده است که دوپامین و همچنین برخی دیگر از تنظیم کننده های عصبی، بر انعطاف پذیری سیناپسی تاثیر می گذارند، مانند تغییر قطبیت و تغییر اندازه پنجره STDP.

یکی از ایده هایی که برای مدلسازی نقش سیستم پاداش به خوبی مطالعه شده است، تعدیل و یا حتی معکوس کردن تغییر وزن تعیین شده توسط STEP است که R-STDP نامیده می شود. R-STDP رد سیناپس های واجد شرایط STDP را ذخیره می کند و تغییرات وزن را در زمان دریافت سیگنال (پاداش یا تنبیه) اعمال می کند.

STDP علاقه به استخراج ویژگی های تکرار شونده دارد که لزوما برای کار مورد نظر مناسب نیستند.

R-STDP بازده محاسباتی را افزایش می دهد.

۲. قانون یادگیری R-STDP

می خواهیم STDP را به سمتی ببریم که تحت تاثیر دوپامین مغز قرار بگیرد.

s وزن سیناپسی است و c نشان دهنده تاثیر پردازش های آهسته ای است که توسط ژن ها و مکانیزم مولکولی صورت می گیرد و زمانی که activity داریم تغییر می کند. در واقع c (eligibility trace) مشخص می کند که آیا اخیرا از سیناپس ما پالسی رد شده است یا خیر.

$$\frac{dc}{dt} = -\frac{c}{\tau_c} + STDP(\tau)\delta(t - t_{pre/post})$$

$-\frac{c}{\tau_c}$ ترمی که decay بشه یعنی وقتی هیچ اتفاقی رخ نمی دهد به آرامی به حالت صفر خود باز گردد. ثابت زمانی نیز در اینجا مقدار نسبتا بزرگی دارد (در اردر ۱ ثانیه) که زمانی که آنزیم ها در حال فعالیت هستند، بازه چند ثانیه ای به طول بینجامد و لحظه ای نباشد. لحظاتی که نورون pre/post synaptic اسپایک می زند، مقدار c به اندازه STDP مربوط به آن اختلاف زمانی، تغییر می کند. درSTDP زمانی که اول نورون presynaptic

اسپایک میزد، LTP و در حالت برعکس، LTD داشتیم. لحظاتی که نورن pre/post synaptic در آن اسپایک می زنند، مقدار تابع δ برابر با ۱ می شود و به اندازه $STDP(\tau)$ ، c افزایش یا کاهش می یابد.

$$\frac{ds}{dt} = cd$$

d نشان دهنده غلظت دوپامین در مغز می باشد. زمانی که دوپامین ترشح می شود نیز با یک نرخی decay می شود و درجا از بین نمی رود. اگر d مثبت باشد، STDP عادی رخ می دهد و اگر منفی باشد، anti STDP رخ می دهد.

$$\frac{dd}{dt} = -\frac{d}{\tau_d} + DA(t)$$

$DA(t)$ بیانگر میزان دوپامین ترشح شده در لحظه t می باشد.

آنچه میزان ترشح دوپامین را مشخص می کند، تفاضل reward و expected reward است.

۳. منابع

Bing, Z., Meschede, C., Chen, G., Knoll, A., & Huang, K. (2020). Indirect and direct training of spiking neural networks for end-to-end control of a lane-keeping vehicle. *Neural Networks*, *121*, 21-36.

Mozafari, M., Kheradpisheh, S. R., Masquelier, T., Nowzari-Dalini, A., & Ganjtabesh, M. (2018). First-spike-based visual categorization using reward-modulated STDP. *IEEE transactions on neural networks and learning systems*, *29*(12), 6178-6190.