Dizajn nového učiaceho pravidla pre moderné Hopfieldove siete

Matej Fandl, Martin Takáč

Centrum pre kognitívnu vedu FMFI
Univerzita Komenského v Bratislave
{matej.fandl,martin.takac}@fmph.uniba.sk

Abstrakt

Asociatívna pamäť je vo výpočtových systémoch využiteľná na rekonštrukciu nekompletných, zašumených vzorov. Moderné Hopfieldove siete sú jej neurálny model, ktorý má veľkú kapacitu a je použiteľný samostatne, ale aj ako súčasť architektúr hlbokého učenia. V našej práci skúmame nové spôsoby učenia týchto sietí. Tento príspevok je vhľadom do prebiehajúcej exploratívnej fázy tvorby nového učiaceho pravidla s ambíciou kombinovať efektívnosť učenia s efektívnosťou samotnej rekonštrukcie.

1 Úvod

Moderné Hopfieldove siete (Krotov a Hopfield, 2016; Ramsauer a spol., 2020) sú trieda modelov asociatívnej pamäte s vysokou kapacitou. Vidíme príležitosť pre efektívne trénovanie týchto sietí tvorbou receptívnych polí neurónov na skrytej vrstve tak, aby využili pravidelnosti v trénovacej množine a rozdelili si prácu. Biologicky inšpirované metódy učenia sa receptívnych polí bez učiteľa vzorov opísali napríklad Krotov a Hopfield (2019), alebo Ravichandran a spol. (2021), nie však pre potreby rekonštrukcie vzorov, ktoré sú oproti klasifikácii špecifické v tom, že ich kombináciou musia vznikať kompletné vzory. Náš výskum sa zameriava na tvorbu nového učiaceho pravidla pre spojité moderné Hopfieldove siete (SMHS) vhodného pre rekonštrukciu.

2 Model

Neuróny v SMHS majú spontánnu aktivačnú dynamiku určenú pravidlom

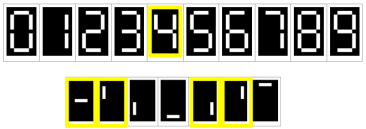
$$\boldsymbol{\xi}^{t+1} = \boldsymbol{W} softmax(\beta \boldsymbol{W}^T \boldsymbol{\xi}^t), \qquad (1)$$

kde $\boldsymbol{\xi}$ je vektor stavov neurónov o dĺžke rovnej dimenzionalite vstupného priestoru d, $\boldsymbol{W}$ je matica s rozmermi $d \times N$, ktorej N stĺpcov predstavuje jednotlivé zapamätané reprezentácie a β je inverzná teplota.

SMHS je s týmto rekonštrukčným pravidlom interpretovateľná ako sieť s jednou skrytou vrstvou (Krotov a Hopfield, 2020). Rekonštruované vzory sú afinnou kombináciou váhových vektorov (receptívnych polí) neurónov skrytej vrstvy.

3 Trénovanie

Potrebu nového spôsobu trénovania spojitých moderných Hopfieldových sietí sme popísali v našej predchádzajúcej práci (Fandl a Takáč, 2022). V nej sme dosiahli deľbu práce neurónov na skrytej vrstve, ale nie požadované distribuované reprezentácie. V tejto sekcii predstavíme nový prístup, ktorým ich chceme dosiahnuť.



Obr. 1: Dátová množina digitálnych číslic a predstava o možnej podobe distribuovaných reprezentácií. Zdroj obrázka: (Fandl a Takáč, 2022)

3.1 Doúčanie sa chýbajúcich častí

Nové pravidlo je postavené na princípe doúčania sa chýbajúcich častí v rekonštrukcii práve predkladaného vzoru. Doúčanie sa je navrhnuté vzhľadom na aktivačnú dynamiku v rovnici (1) a jeho fungovanie opisuje pseudokód jednej epochy.

for all $vzor \leftarrow vzory$ **do**
  $rez = kopiruj(vzor)$
  while $\|rez\| > \epsilon$ **or** $!max\_iteracii$ **do**
    $vaha = najblizsia\_k\_rez$
    $vaha = \frac{vaha + \alpha \cdot rez}{\|vaha + \alpha \cdot rez\|}$
    $rez = rez - p_{vaha} \cdot vaha$
  end while
end for

V algoritme je ϵ prahová hodnota zanedbateľnosti rezídua, α je rýchlosť učenia a p_{vaha} je miera pravdepodobnosti, do akej receptívne pole reprezentuje rezíduum. Táto hodnota je získaná ako príslušný komponent vektora $\boldsymbol{p} = softmax(\beta \boldsymbol{W}^T \boldsymbol{rez})$, kde $\boldsymbol{W}$ je matica váh SMHS. Vzorec pre výpočet $\boldsymbol{p}$ je adaptovaný z rovnice (1).

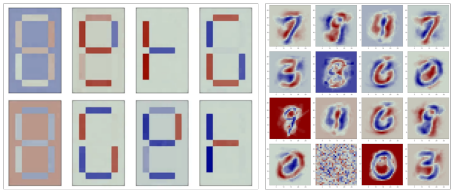
4 Predbežné výsledky

Prezentované výsledky exploratívnej fázy sú z experimentov na jednoduchej dátovej množine 13 $\times$ 8 px

digitálnych číslic (DČ) a na dátovej množine MNIST (Deng, 2012).

4.1 Receptívne polia

Podobne ako pri predchádzajúcom prístupe (Fandl a Takáč, 2022) sa nám darí dosiahnuť deľbu práce medzi neurónmi, avšak tentoraz aj s distribuovanými reprezentáciami. Pre obe množiny sú výsledné receptívne polia vidieť na Obr. 2.



Obr. 2: Receptívne polia pri trénovaní na dátových množinách DČ a MNIST. Červená - negatívne hodnoty, modrá - pozitívne hodnoty, biela - hodnoty blízke nule.

V oboch prípadoch sa tvorí „priemerný neurón“, ktorého váhový vektor je blízko ku všetkým vzorom, a páry „opačných neurónov“, ktoré sa vzájomne vylučujú.

Tieto páry sú spôsobené osciláciami zapríčinenými odpočítavaním normalizovaných váhových vektorov od nenormalizovaného rezídua pri povolenom opakovanom výbere neurónov. Biele časti receptívnych polí slúžia vďaka použitej metrike – kosínusovej podobnosti – ako maska. Väčšina neurónov teda slúži ako detektory príznakov.

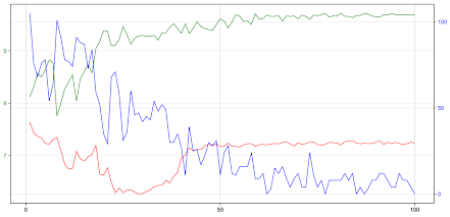
4.2 Kvalita rekonštrukcie

Meriame rozdiel medzi vstupom a rekonštrukciou kompletných, nezašumených vzorov. Konkrétne kumulatívnu kosínusovú podobnosť, euklidovskú vzdialenosť a počet komponentov s odlišným znamienkom. Pre dátovú množinu DČ sa nám podarilo dosiahnuť nulovú chybu pri poslednej metrike, avšak spojitá rekonštrukcia je neuspokojivá. Pre MNIST sa nám zatiaľ nepodarilo dosiahnuť uspokojivé výsledky v žiadnej zo spomínaných metrík.

Na Obr. 3 je vidieť úspešný beh. Za pozornosť stojí zhoršovanie rekonštrukcie po úspešnej 60. epoche.

5 Záver

Predbežné výsledky naznačujú, že pokračovať v skúmaní pravidla má zmysel, ale aj poukazujú na nedostatky a otvorené otázky. Dosahujeme deľbu práce medzi neurónmi aj distribuované reprezentácie. Lepšie porozumenie vplyvu parametrov a stratégie selekcie váhových vektorov na dynamiku učiaceho pravidla by mohlo viesť k jeho potrebným úpravám pre zlepšenie



Obr. 3: Vývoj rekonštrukčnej chyby v epochách trénovania na množine digitálnych číslic. Zelená - kosínová podobnosť, červená - euklidovská vzdialenosť, modrá - počet komponentov s odlišným znamienkom.

kvality rekonštrukcie. Receptívne polia formované trénovaním na dátovej množine MNIST pôsobia podobne ako tie, ktoré dosiahli Krotov a Hopfield (2019). Jedným z plánov je otestovať vhodnosť tých našich aj pri klasifikácii.

Poďakovanie

Tento príspevok vznikol s podporou grantu VEGA 1/0373/23 a KEGA 022UK-4/2023.

Literatúra

Deng, L. (2012). The MNIST database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 29(6):141–142.

Fandl, M. a Takáč, M. (2022). Zvyšovanie efektivity trénovania a kapacity v atraktorovom neurálnom modeli asociatívnej pamäte. V *Kognice a umělý život XX*.

Krotov, D. a Hopfield, J. (2020). Large associative memory problem in neurobiology and machine learning. *arXiv preprint arXiv:2008.06996*.

Krotov, D. a Hopfield, J. J. (2016). Dense associative memory for pattern recognition. *Advances in Neural Information Processing Systems 29 (2016), 1172-1180*.

Krotov, D. a Hopfield, J. J. (2019). Unsupervised learning by competing hidden units. *Proceedings of the National Academy of Sciences of the United States of America*, 116(16):7723–7731.

Ramsauer, H. a spol. (2020). Hopfield networks is all you need. *arXiv preprint arXiv:2008.02217*.

Ravichandran, N. B., Lansner, A. a Herman, P. (2021). Brain-like approaches to unsupervised learning of hidden representations – a comparative study. Farkaš, I. a spol. (zost.), V *Artificial Neural Networks and Machine Learning - ICANN 2021*, str. 162–173, Cham. Springer International Publishing.