

Hopfield Networks : Attention is all you need

Валитов Эльдар, БПМИ-172



Oct 14, 2020

Классическая сеть Хопфилда



Нейронная сеть Хопфилда(1982)

- ▶ Полносвязная нейронная сеть с симметричной матрицей весов. Матрица весов W хранит паттерны ξ

$$W = \sum_i^N x_i x_i^T . \quad x_i \in \{-1, 1\}^d$$

Синхронное правило обновления

$$\xi^{t+1} = \text{sgn}(W\xi^t - b)$$

Асинхронное правило обновления

$$E = -\frac{1}{2}\boldsymbol{\xi}^T \mathbf{W} \boldsymbol{\xi} + \boldsymbol{\xi}^T \mathbf{b} = -\frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d w_{ij} \xi_i \xi_j + \sum_{i=1}^d b_i \xi_i$$

Пример работы



$$W = \mathbf{x}_{\text{Homer}} \mathbf{x}_{\text{Homer}}^T, \quad \mathbf{x}_{\text{Homer}} \in \{-1, 1\}^d$$

Пример работы



Проблемы

$$W = \sum_{i=1}^3 x_i x_i^T, \quad x_i \in \{-1, 1\}^d.$$



train input 1



train input 2



train input 3



masked test image



retrieved



train input 1



train input 2



train input 3



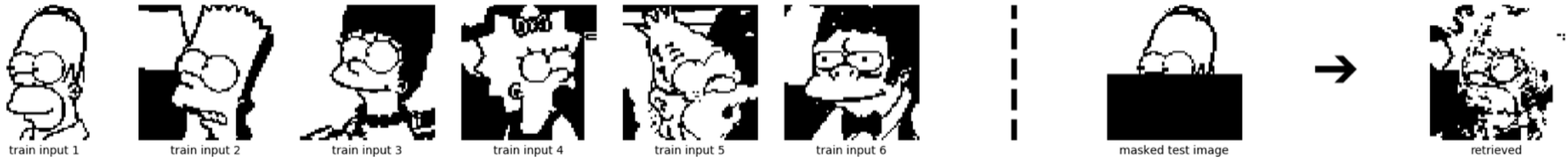
masked test image



retrieved

Память

$$C \cong \frac{d}{2 \log(d)}, \text{ где } d - \text{размерность входа}$$



$$C \cong 0.14d = 0.14 \cdot 64 \cdot 64 \sim 570$$

Проблемы

- ▶ Относительно небольшая память для хранения паттернов
- ▶ Бинарность значений
- ▶ Для визуально близких изображений восстановление затруднительно

Современная сеть Хопфилда



Современная сеть Хопфилда

Krotov и Hopfield:
$$E = - \sum_{i=1}^N F(x_i^T \xi) \quad F(z) = z^a$$

Память:
$$C \cong \frac{1}{2(2a-3)!!} \frac{d^{a-1}}{\log(d)}$$

Память нежесткая:
$$C \cong \alpha_a d^{a-1}$$

Современная сеть Хопфилда

Demircigil:

$$E = - \sum_{i=1}^N \exp(\mathbf{x}_i^T \boldsymbol{\xi}) \quad F(z) = \exp(z)$$

$$E = -\exp(\text{lse}(1, \mathbf{X}^T \boldsymbol{\xi}))$$

$$\text{lse}(\beta, \mathbf{z}) = \beta^{-1} \log \left(\sum_{l=1}^N \exp(\beta z_l) \right)$$

Новая память:

$$C \cong 2^{\frac{d}{2}}$$

Новое правило обновления

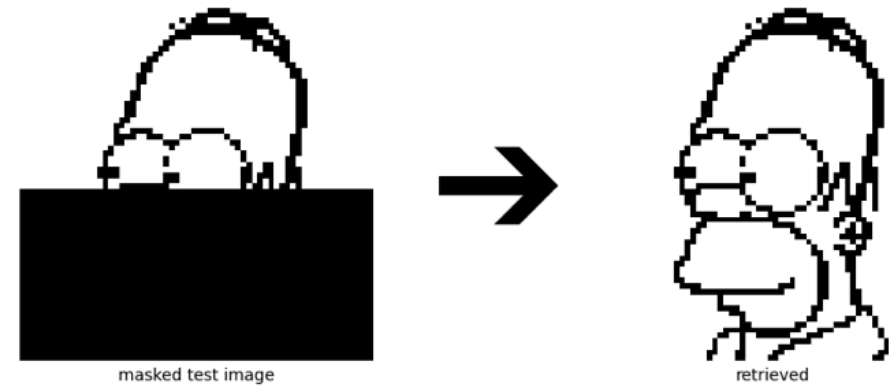
$$\xi^{\text{new}}[l] = \text{sgn} \left[-E(\xi^{(l+)}) + E(\xi^{(l-)}) \right]$$

$$\xi^{\text{new}}[l] = \text{sgn} \left[\sum_{i=1}^N \exp(x_i^T \xi^{(l+)}) - \sum_{i=1}^N \exp(x_i^T \xi^{(l-)}) \right]$$

$$\xi^{(l+)}[l] = 1 \text{ and } \xi^{(l-)}[l] = -1 \text{ and } \xi^{(l+)}[k] = \xi^{(l-)}[k] = \xi[k] \text{ for } k \neq l.$$

Согласно статье Demircigil, с высокой вероятностью сходится за один асинхронный шаг

Пример работы



Новая функция энергии. Связь с
Attention.



Новая функция энергии

$$E = -\text{lse}(\beta, X^T \xi) + \frac{1}{2} \xi^T \xi + \beta^{-1} \log N + \frac{1}{2} M^2$$

Позволяет хранить непрерывные паттерны

CCCP(Concave-Convex-Procedure)

- the total energy $E(\xi)$ is split into a convex and a concave term: $E(\xi) = E_1(\xi) + E_2(\xi)$
- the term $\frac{1}{2}\xi^T\xi + C = E_1(\xi)$ is convex (C is a constant independent of ξ)
- the term $-\text{lse}(\beta, X^T\xi) = E_2(\xi)$ is concave (lse is convex since its Hessian is positive semi-definite, which is shown in the appendix of the paper)
- The CCCP applied to E is:

$$\nabla_{\xi} E_1(\xi^{t+1}) = -\nabla_{\xi} E_2(\xi^t)$$

$$\nabla_{\xi} \left(\frac{1}{2}\xi^T\xi + C \right) (\xi^{t+1}) = \nabla_{\xi} \text{lse}(\beta, X^T\xi^t)$$

$$\xi^{t+1} = X \text{softmax}(\beta X^T\xi^t) ,$$

где $\nabla_{\xi} \text{lse}(\beta, X^T\xi) = X \text{softmax}(\beta X^T\xi)$

СССР(Concave-Convex-Procedure)

Правило обновления теперь выглядит так:

$$\xi^{\text{new}} = X \text{softmax}(\beta X^T \xi)$$

Основные свойства нашей новой функции энергии:

- Глобальная сходимость к локальному минимуму
- Экспоненциальная память
- Сходимость после одного шага обновления
- Применение СССР гарантирует монотонное уменьшение функции энергии

Пример работы



train input 1



train input 2



train input 3



train input 4



train input 5



train input 6



train input 7



train input 8



train input 9



train input 10



train input 11



train input 12



train input 13



train input 14



train input 15



train input 16



train input 17



train input 18



train input 19



train input 20



train input 21



train input 22



train input 23



train input 24

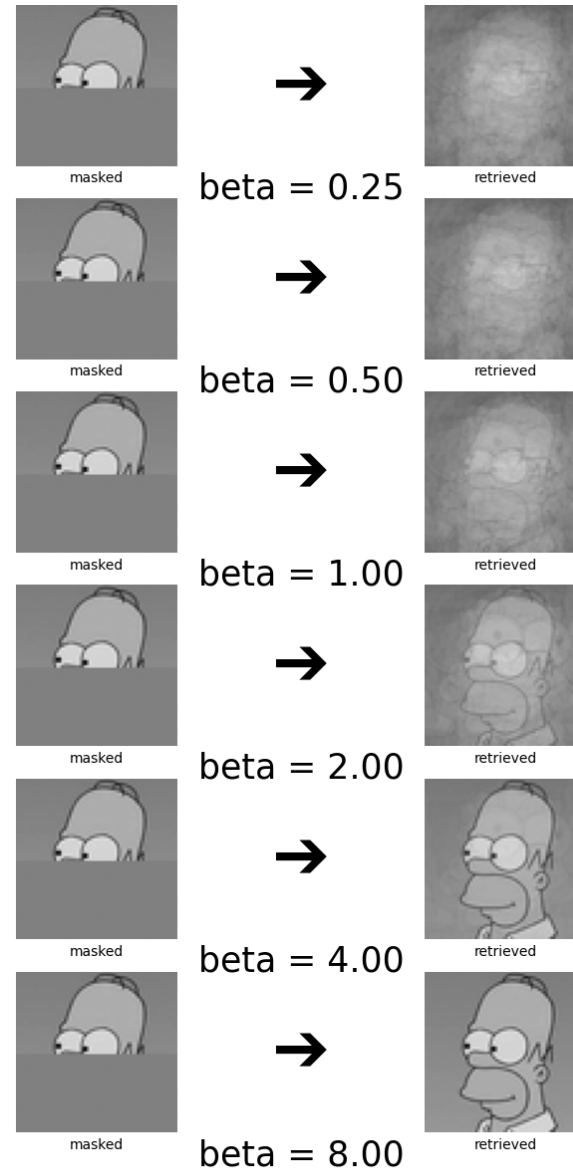


masked test image

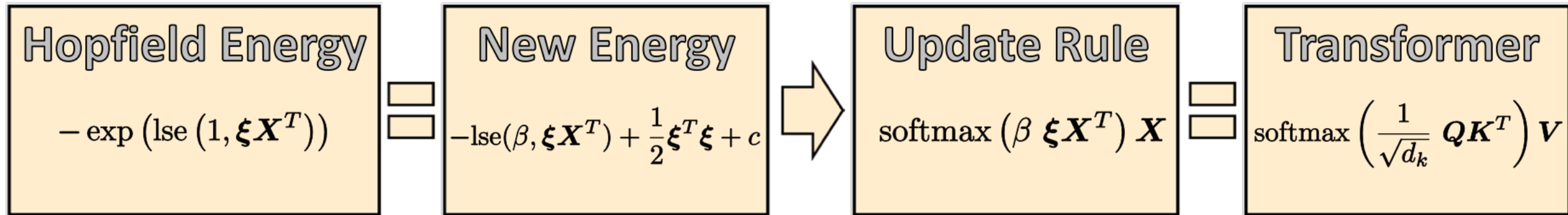


retrieved

Пример работы



Связь с Attention



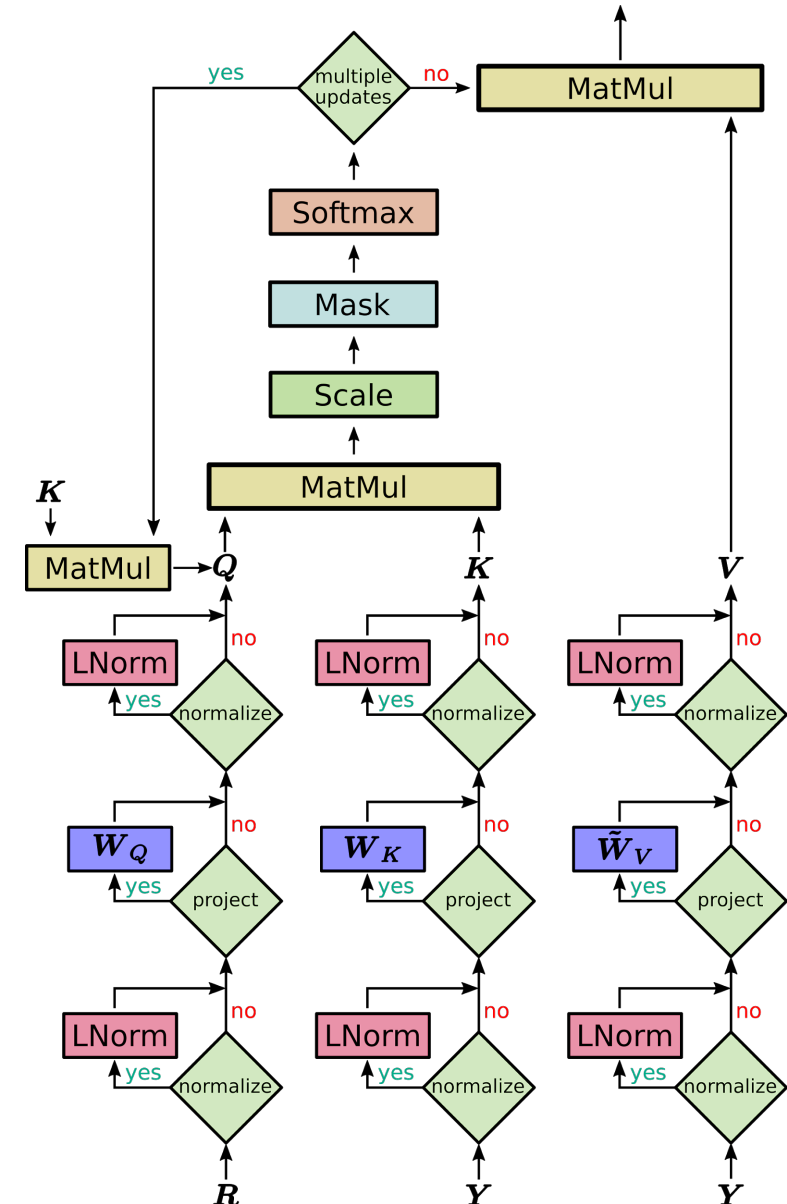
Практическое применение.



Описание слоя PyTorch

$$Z = \text{softmax}(\beta \cdot RW_Q W_K^T Y^T) Y W_K W_V$$

- **Association of two sets**
- **Variable** β that determines the kind of fixed points
- **Multiple updates** for precise fixed points
- **Dimension of the associative space** for controlling the storage capacity
- **Static patterns** for fixed pattern search
- **Pattern normalization** to control the fixed point dynamics by norm and shift of the patterns



Описание слоя PyTorch

$$\mathbf{Z} = \text{softmax} \left(\beta \mathbf{R} \mathbf{W}_Q \mathbf{W}_K^T \mathbf{Y}^T \right) \mathbf{Y} \mathbf{W}_K \mathbf{W}_V$$

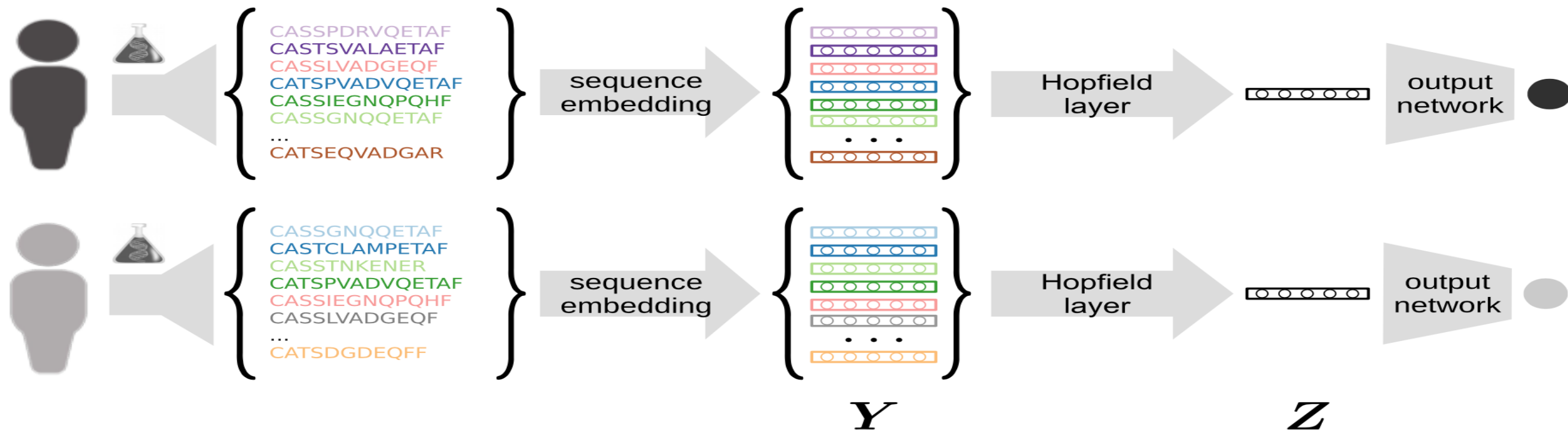
Практический пример:

$$\mathbf{Z} = \text{softmax} \left(\beta \mathbf{R} \mathbf{Y}^T \right) \mathbf{Y}$$

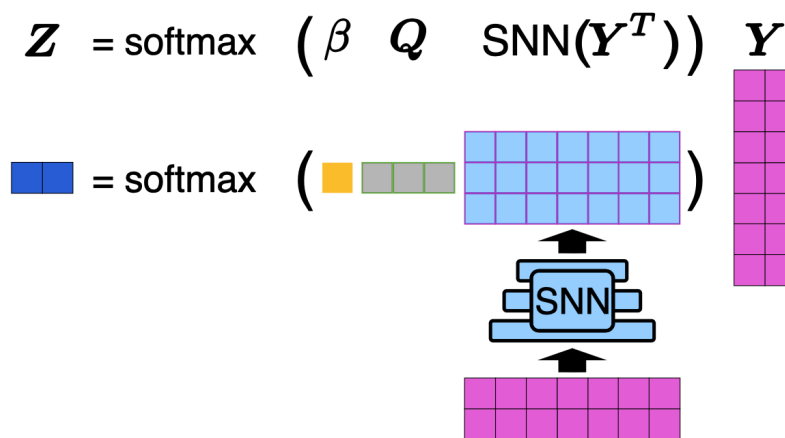
Описание слоя PyTorch (Pooling)

$$\begin{array}{c}
 \mathbf{Z} = \text{softmax} \left(\beta \quad \mathbf{Q} \quad \mathbf{W}_K^T \mathbf{Y}^T \right) \quad \mathbf{Y} \quad \mathbf{W}_K \quad \mathbf{W}_V \\
 \text{[4 blue boxes]} = \text{softmax} \left(\text{[1 yellow box]} \quad \text{[3 green boxes]} \quad \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \right) \quad \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} \quad \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} \quad \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array}
 \end{array}$$

DeepRC



Hopfield layer:



BERT



Вопросы

- ▶ Какие проблемы существуют у классической сети Хопфилда и как их решает современная версия?
- ▶ Какая аналогия между новой функцией энергии и механизмом attention?
- ▶ Что из себя представляет классическая сеть Хопфилда, какую задачу она решает и каким образом?

Источники

- ▶ https://ml-jku.github.io/hopfield-layers/#mjx-eqn-equpdate_sepp4 - блог о статье
- ▶ <https://arxiv.org/pdf/2008.02217.pdf> - статья
- ▶ <https://www.youtube.com/watch?v=nv6oFDp6rNQ> - видео разбор статьи