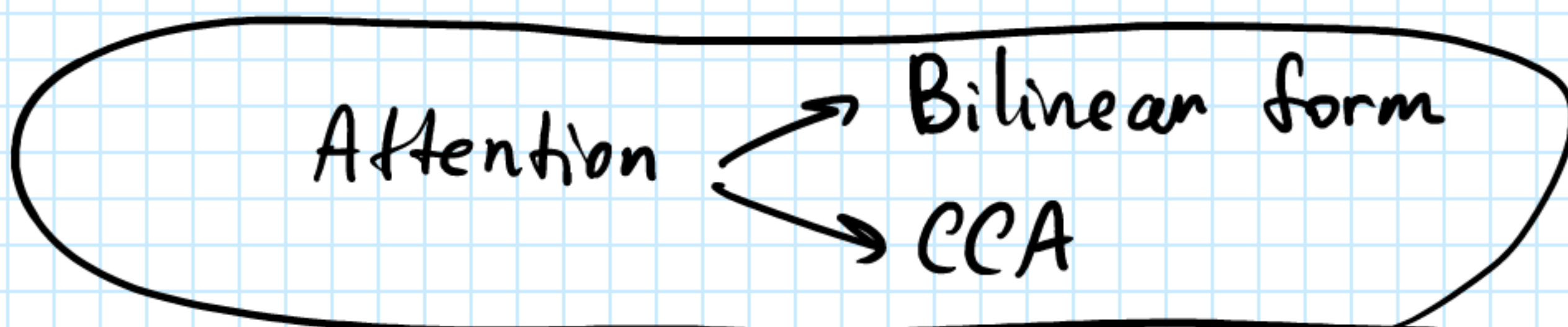


Seq-2-Seq, SSM, RNN-LSTM with cross-attention

$$f: \underset{\text{EEG}}{X} \rightarrow \underset{\text{IMU}}{Y}$$



Transformer version (BERT) Энумерация

Causality in time: $\left[\begin{matrix} \text{Riem. Space} \\ \text{PINN} \end{matrix} \rightarrow \begin{matrix} \text{Hamilt. space} \end{matrix} \right]$

Attention

$$\text{self-attn}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right) V$$

$T \times d \quad T \times d \quad T \times d$ $T \times T$ $T \times d$

Multihead Attn

$$\text{MHAttn}(Q, K, V) = [\text{head}_1 \dots \text{head}_p] W^Q$$

$T \times d$ $T \times pd$ $pd \times d$

$$\text{head}_i = \text{SelfAttn}(QW_i^Q, KW_i^K, VW_i^V)$$

$$\text{self-attn}(X) = \text{self-attn}(XW_k, XW_q, XW_v)$$

$T \times m$ $T \times m \times d$ $T \times m \times d$

$$\text{self-attn}: \mathbb{R}^{T \times m} \rightarrow \mathbb{R}^{T \times d}$$

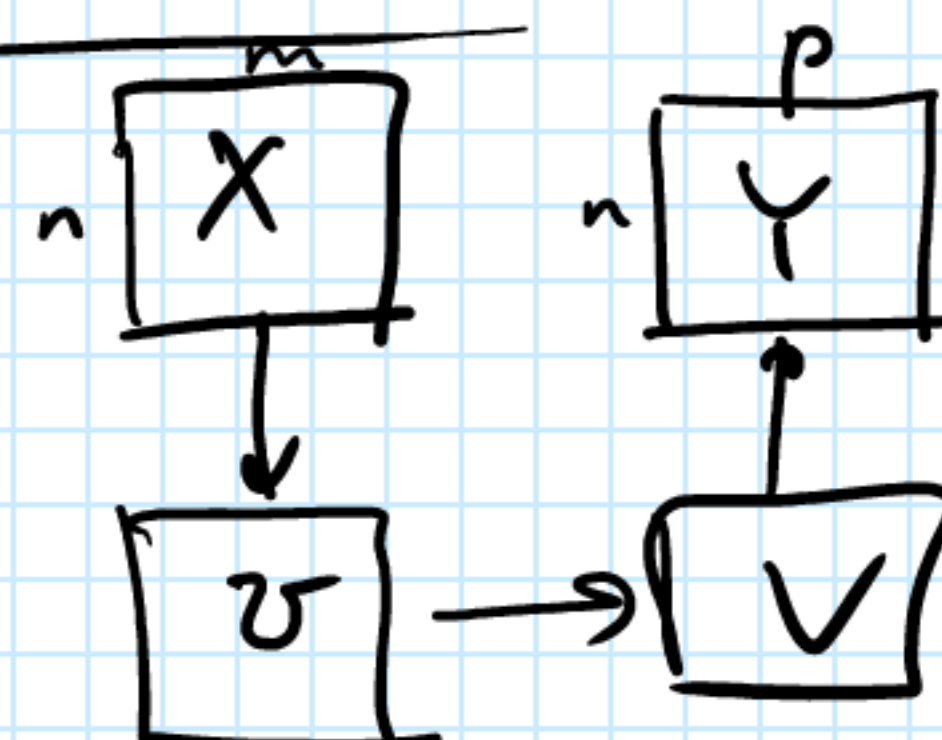
$$\text{Cross-attention: } \text{cross-attn}(X_1, X_2) = \text{self-attn}(X_1W_q, X_2W_k, X_1W_v)$$

from decoder from encoder

$$\max J = W^T X^T Y c$$

s.t. $\|XW\| = 1$ $1 \times m \quad m \times n \quad n \times p \quad p \times 1$

$\|Yc\| = 1$



$$\max g(x, y) = \frac{E[W_x X^T Y W_y^T]}{\sqrt{E[W_x X^T X W_x^T]} \sqrt{E[W_y Y^T Y W_y^T]}}$$

self-attn

$$\varphi_w: \mathbb{R}^{d \times t} \rightarrow \mathbb{R}^{d' \times t}$$

$$Q, K \in \mathbb{R}^{a \times d}$$

$$(x_i)_{1 \leq i \leq t} \mapsto \left(\sum_{1 \leq j \leq t} x_j^T K^T Q x_i V x_j \right)_{1 \leq i \leq t}$$

$1 \times d \quad d \times a \quad a \times d \quad d \times 1 \quad d' \times d \quad d \times 1$

$$\varphi_w(X) = V X X^T K^T Q X$$

$d' \times d \quad d \times t \quad t \times d \quad d \times d \quad d \times t \quad d' \times t$

$X^T A y$ - bilinear form

$$\text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right) V$$

$$\text{softmax}\left(\frac{XW_q W_k^T X^T}{\sqrt{d}}\right) XW_v \in \mathbb{R}^{t \times d'}$$

$t \times d \quad d \times a \quad a \times d \quad d \times n \quad t \times d \quad d' \times d'$

$$O(d^2 t + d^2 t + d^3 + d^2 t) = O(d^2 t + d^3)$$

$$X Q K^T X^T X V$$

$t \times d \quad d \times d \quad d \times t \quad t \times d \quad d \times d'$

$O(t^2 d + d^3)$

Уменьшение асимптотики за счёт удадения нормализации и использование ассоциативности

$$f: \mathbb{R} \rightarrow \mathbb{R}_{++}$$

$$f(0) = 1, f - \text{инкрементна}$$

$$\text{Example: } f(x) = e^{x/2}$$

$$\varphi_w(X) = \left(\frac{1}{Z_i} \sum_{1 \leq j \leq t} \delta(x_i^T A x_j) V x_j \right)_{1 \leq i \leq t} \quad Z_i = \sum_{1 \leq k \leq t} \delta(x_i^T A x_k)$$

$$(w_1^*, w_2^*) = \underset{\substack{w_1 \in \mathbb{R}^{n_1} \\ w_2 \in \mathbb{R}^{n_2}}}{\text{argmax}} \text{corr}(w_1^T X_1, w_2^T X_2) = \underset{\substack{w_1 \in \mathbb{R}^{n_1} \\ w_2 \in \mathbb{R}^{n_2}}}{\text{argmax}} \frac{w_1^T \Sigma_{12} w_2}{\sqrt{w_1^T \Sigma_{11} w_1} \sqrt{w_2^T \Sigma_{22} w_2}}$$

$$X_1 \in \mathbb{R}^{n_1}, X_2 \in \mathbb{R}^{n_2}$$

Top $k \leq \min(n_1, n_2)$ projections:

$$\max \text{tr}(A_1^T \Sigma_{12} A_2)$$

s.t. $A_1^T \Sigma_{11} A_1 = I$
 $A_2^T \Sigma_{22} A_2 = I$

$$T = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} - \text{combination of cross-attention with "modified" self-attention}$$

 U_k, V_k - first k left and right singular vectors

$$\text{Optimal} = \sum_{j=1}^k \sigma_j \text{ (from } T \text{)}$$

$$(A_1^*, A_2^*) = (\Sigma_{11}^{-1/2} U_k, \Sigma_{22}^{-1/2} V_k)$$

$$\text{DL + CCA: } (u_1^*, u_2^*, w_1^*, w_2^*) = \underset{\substack{u_1, u_2 \\ X_i \in \mathbb{R}^{n \times d_i}}}{\text{argmax}} \text{corr}(u_1^T f_{w_1}(X_1), u_2^T f_{w_2}(X_2))$$