

Input

$$P_m: e_1, e_2, \dots, e_n$$

Trainable B-spline Functions
($KAN_{transform}$)

$$\phi: \text{[B-spline functions]}$$

$$H_m: \phi(e_1), \phi(e_2), \dots, \phi(e_n)$$

$$H_m \rightarrow \text{Concat}(H_m \oplus \mathbf{1}, \mathbf{1} \oplus H_m)$$

Trainable B-spline Functions (KAN_m)

$$\text{[B-spline functions]}$$

The Mapping Function f

$$\text{[B-spline functions]}$$

$$a_t = [a_{t,1}, \dots, a_{t,2}, \dots, a_{t,n}]$$

• Matrix Multiplication

KAN-Attention:

$$\text{Input: } P_m = \{e_1^m, e_2^m, e_3^m, \dots, e_n^m\}, m = \{T, A, V\}$$

$$\begin{aligned} \text{Matrix: } H_T &= [\phi_Q(e_1^m), \phi_Q(e_2^m), \phi_Q(e_3^m), \dots, \phi_Q(e_n^m)] \\ H_A &= [\phi_K(e_1^m), \phi_K(e_2^m), \phi_K(e_3^m), \dots, \phi_K(e_n^m)] \\ H_V &= [\phi_V(e_1^m), \phi_V(e_2^m), \phi_V(e_3^m), \dots, \phi_V(e_n^m)] \end{aligned}$$

$$\text{Get: } H_T, H_A, H_V$$

Equation 6

Attention Weight:

$$a_t = KAN_t(H_T)a_a = KAN_a(H_A)a_v = KAN_v(H_V)$$

Equation 7

$$\alpha_F = \mathbf{w}(\alpha_t + \alpha_a + \alpha_v) + \text{bias}$$

$$\text{Get: } \alpha_F$$

Equation 10

$$\begin{aligned} P'_T &= KAN(\text{Softmax}(\alpha_F) \cdot H_T) \\ P'_A &= KAN(\text{Softmax}(\alpha_F) \cdot H_A) \\ P'_V &= KAN(\text{Softmax}(\alpha_F) \cdot H_V) \end{aligned}$$

Equation 11,12,13

$$\text{Output: } P'_T, P'_A, P'_V$$

$$\begin{aligned} &\underbrace{\alpha_t}_{\omega_t} + \underbrace{\alpha_a}_{\omega_a} + \underbrace{\alpha_v}_{\omega_v} \\ &\alpha_F = \text{[B-spline functions]} + \text{bias} \end{aligned}$$

$$\begin{aligned} &\alpha_F \text{ [B-spline functions]} \quad V_T \text{ [B-spline functions]} \\ &\text{Kolmogorov-Arnold Networks} \\ &P'_t \text{ [B-spline functions]} \end{aligned}$$