$Q \in R^{L_Q \times d_{model}}$ $K \in R^{L_k \times d_{model}}$

$\frac{1}{\sqrt{d_k}} \times sum(\quad \cdot [\quad]^T)$

$\mathcal{A}(Q,K,V) = \frac{QK^T}{\sqrt{d_k}} V$

(a) Attention

Q

$\overline{K} \in R^{L_{\overline{K}} \times d_{model}}$

$\overline{Q} \in R^{L_{\overline{Q}} \times d_{model}}$

$\overline{Q}$ K

$Top(\quad \cdot [\quad]^T) \rightarrow \quad \rightarrow \frac{1}{\sqrt{d_k}} \times sum(\quad \cdot \quad)$

$\mathcal{A}(Q,K,V) = \frac{\overline{Q} K^T}{\sqrt{d_k}} V$

(b) Probsparse attention

$Q_{period} \in R^{L_{period} \times d_{model}}$

$\overline{K}$ $A_h \in R^{L_{\overline{K}} \times d_{model}}$

$\overline{Q}$

$K_{period} \in R^{L_{period} \times d_{model}}$

$\overline{Q}$

$Top (\quad \cdot [\quad \odot \quad]^T) \rightarrow \quad \rightarrow Expand(\frac{1}{\sqrt{d_k}} \times sum[\quad \cdot \quad]^T)$

$\mathcal{A}(Q,K,V) = \frac{\overline{Q} K^T}{\sqrt{d_k}} V$

(c) Periodic attention

$\cdot$ Matrix product

$\odot$ Hadamard product

Time steps

Random selection of time steps

Time steps after sorting

Time steps selected by the first periodic term

Time steps selected by second periodic term

Time steps selected by local range

Amplitude