Methodology
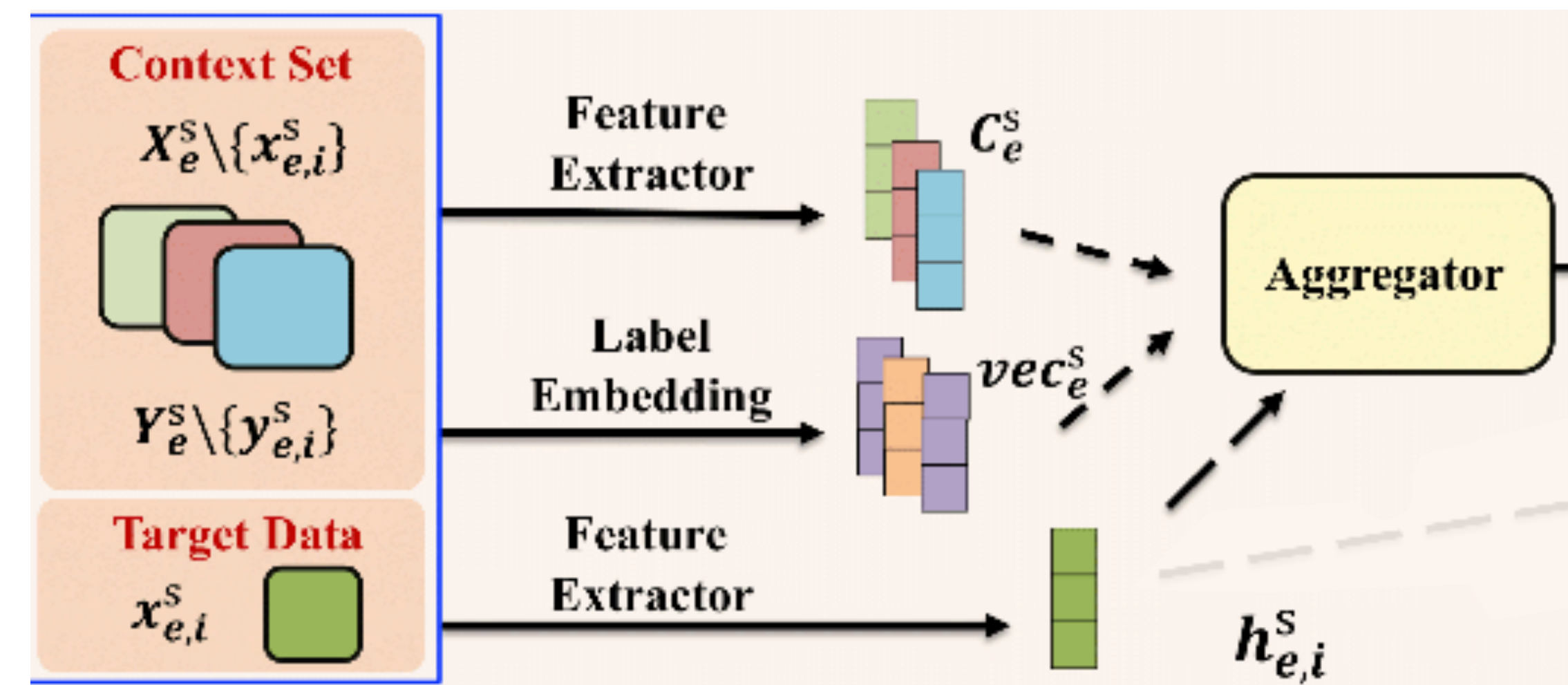## Aggregator

- Design aggregator satisfies two properties:

  - Permutation-invariant & Target-dependent

  - Adopt the attention mechanism

    - Compute weights of each observations in context set with respect to the target and aggregate the values according to their weights to form the new value.

Methodology
Aggregator: Attention

- Use scaled dot-product attention mechanism

- Mapping a query $\mathbf{Q}$ and a set of key $\mathbf{K}$ - value $\mathbf{V}$ pairs to an output

- $\mathbf{Q}_i = \mathbf{W}_q \mathbf{h}_{e,i}$, $\mathbf{K} = \mathbf{W}_k \mathbf{C}_e$, $\mathbf{V} = \mathbf{W}_v \left(\mathbf{C}_e \oplus \mathbf{vec}_e \right)$

- $a_i = \text{softmax} \left(\dfrac{\mathbf{Q}_i \, \mathrm{K}^T}{\sqrt{d}} \right)$

- $\text{Attention} \left(\mathbf{Q}_i, \mathbf{K}, \mathbf{V} \right) := a_i \mathbf{V}$