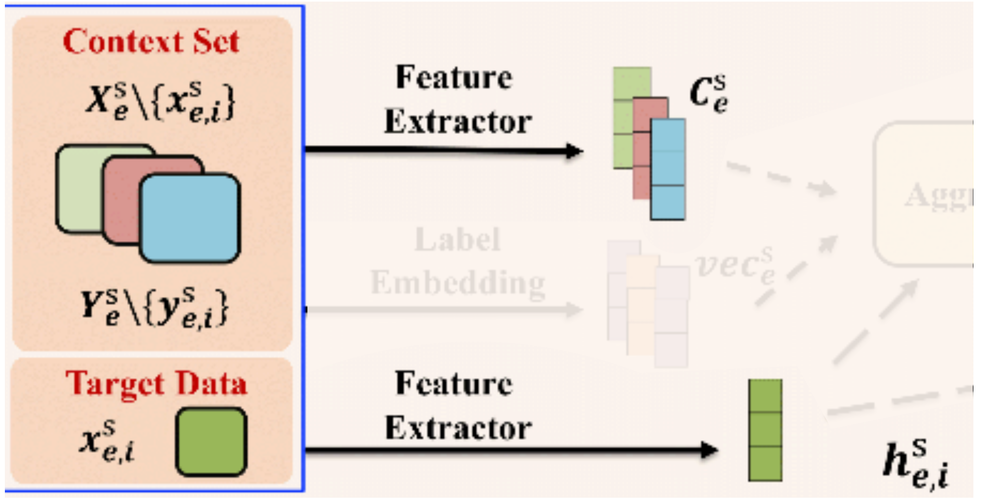
Methodology

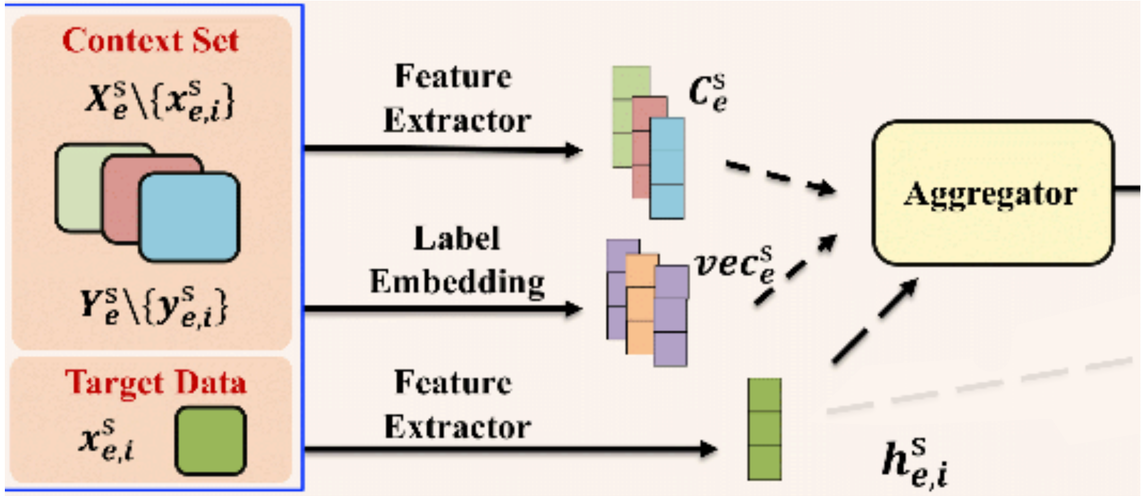
Feature Extractor



- **Textual feature extractor**
  - Text-CNN
  - Use 300-dimensional FastText pre-trained word-embedding
- **Visual feature extractor**
  - Pre-trained VGG-19
- On the top of extractors, both add a fully connected layer to adjust dimension to d_f.
- The output of two extractors are concatenated together to form a feature vector.

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.