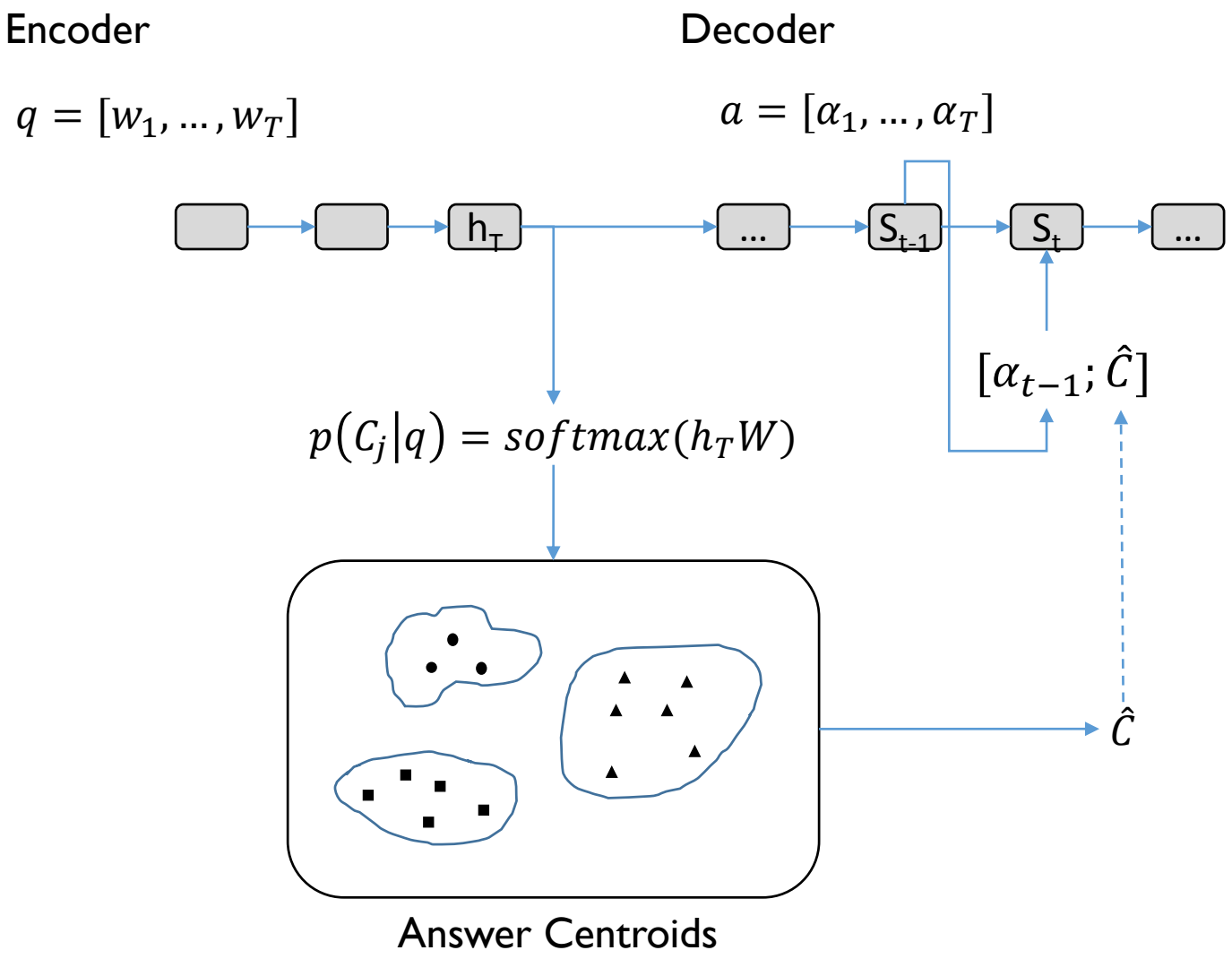

Deep Metric Learning

1. Clustering all answers with their word embeddings (mean of word vectors in each answer sentence)
2. Training Encoder-Decoder model with <q,a,C> triples, where q=question, a=answer and C=answer cluster.
   a) Encoder: an RNN to encode q into a vector h_T
   b) Cluster label predictor: given h_T, predict the potential cluster label $\hat{C}$
   c) Decoder: an RNN to decode answer a according to h_T and answer cluster centroid/label
   d) Train to minimize two objectives
      - Accuracy of cluster prediction
      - Accuracy of decoded answer

$$loss = -\log p(C_j|q) - \sum_{t=1}^{N} \log p(\alpha_t|q, C_j)$$

3. Prediction: given q, first predict C, then decode a using encoded q as well as C.