Algorithm 1 离题检测聚类方案算法

Input:

Number of topics N

Model M obtain context vector

Output:

Recall@10 for each prompts

R for all prompts

for each $i \in [1, N]$ **do**
  Get the context vector $C_j(j = 0, 1, ..., N)$ of all compositions under the prompt P_i by model M;
  Use AgglomerativeClustering algorithm to cluster C_j, get M clusters $R_k(k = 0, 1, ..., M)$
  for each $k \in R_k$ **do**
    if $Num(k)$ >5 **then**
      $Samples$.append(k)
    else
      $MaybeOT$.append(k)
    end if
  end for
  for each $ot \in MaybeOT$ **do**
    Calculate the distance between ot and $Samples$ separately, get $D_j(j = 0, 1, ..., len(Samples))$
    $D = min(D_j)$
  end for
  Sort all D, calculate *Recall*@10, get the $Recall_i$ of P_i
end for
Take the average of all $Recall_i$ to get R
$R = \frac{\sum_{i=0}^{10} Recall_i}{10}$