Pseudo-Code

Algorithm 1: Attribute spatial association based sampling

Input: *A*: the point set; *r*: radius parameter; *min_r*: minimum radius;

Output: *S*: sampled set;

Algorithm: Random: randomly returns an element from the set; KDE: kernel-density estimate; Label: the label of point, processed by the clustering algorithm; Dist: the Euclidean distance between the given two points; Entropy: the entropy of the set;

$S \leftarrow \varnothing$; $T \leftarrow \varnothing$;

while $A \neq \varnothing$ **or** $T \neq \varnothing$ **do**

  if $T \neq \varnothing$ **then**

    $c \leftarrow Random(T)$; $T \leftarrow T \setminus c$; add c into S;

  else

    $c \leftarrow Random(A)$; $A \leftarrow A \setminus c$; add c into S;

  $radius \leftarrow Max(r / KDE(c), min\_r)$;

  $neighbors \leftarrow \varnothing$;

  add $Label(pi)$ into $neighbors$, $Dist(c, pi) \leq radius$, all $pi \in (A + T)$;

  $radius \leftarrow (radius - min\_r) / (1 + Entropy(neighbors)) + min\_r$;

  $disabled \leftarrow \varnothing$; $active \leftarrow \varnothing$;

  add pi into disabled, $Dist(c, pi) \leq radius$, all $pi \in (A + T)$;

  add pi into $active$, $radius < Dist(c, pi) \leq 2 * radius$, all $pi \in (A + T)$;

  $A \leftarrow A \setminus pi$, all $pi \in (disabled + active)$;

  $T \leftarrow T \setminus pi$, all $pi \in (disabled + active)$;

  add pi into T, all $pi \in active$;

end while;

return S;