

Algorithm PROCLUS_SPMMP (No. of clusters: k , Average number of dimensions: l)

$\{C_i$ is the i^{th} cluster}

$\{D_i$ is the set of dimension associated with cluster $C_i\}$

$\{M_{\text{best}}$ is the best set of medoids found so far}

$\{N$ is the final set of medoids with associated dimension}

$\{A, B$ are constant integers}

$\{D$ is the set of all dimension (all the process stages)}

$\{P$ is the set of all points (all the production paths)}

begin

 {1. Initialization Phase}

$S = \text{random sample of size } A \cdot k$

$M = \text{Greedy}(S, B \cdot k)$

 {2. Iterative Phase}

$\text{BestObjective} = \infty$

$M_{\text{current}} = \text{Random set of medoids } \{m_1, m_2, \dots, m_k\} \subset M$

repeat

 {approximate the optimal set of dimensions}

$(L_1, \dots, L_k) = \text{FindLocality}(S, M_{\text{current}})$

$(D_1, D_2, \dots, D_k) = \text{FindDimensions}(k, l, L)$

 {form the clusters}

$(C_1, \dots, C_k) = \text{AssignPoints}(D_1, \dots, D_k)$

$\text{ObjectiveFunction} = \text{EvaluateClusters}(C_1, \dots, C_k, D_1, \dots, D_k)$

if $\text{ObjectiveFunction} < \text{BestObjective}$ **then**

begin

$\text{BestObjective} = \text{ObjectiveFunction}$

$M_{\text{best}} = M_{\text{current}}$

end

 Compute M_{current} by replacing the bad medoids in M_{best} with random points from M

until (termination_criterion)

 {3. Refinement Phase}

$L = \{C_1, \dots, C_k\}$

$(D_1, \dots, D_k) = \text{FindDimensions}(k, l, L)$

$(C_1, \dots, C_k) = \text{AssignPoints}(D_1, \dots, D_k)$

$N = (M_{\text{best}}, D_1, \dots, D_k)$

return(N)

end

Algorithm PROCLUS_SPMMP (No. of clusters: k , Average number of dimensions: l , number of agent T)

$\{C_i$ is the i^{th} cluster}

$\{D_i$ is the set of dimension associated with cluster $C_i\}$

$\{M_{\text{best}}$ is the best set of medoids found so far}

$\{N$ is the final set of medoids with associated dimension}

$\{A, B$ are constant integers}

$\{D$ is the set of all dimension (all the process stages)}

$\{P$ is the set of all points (all the production paths)}

begin

{1. Initialization Phase}

$S = \text{random sample of size } A \cdot k$

$M = \text{Greedy}(S, B \cdot k)$

$\text{BestObjective} = \infty$

for agent $t \in T$ **do**

begin

$t.M_{\text{current}} = \text{Random set of medoids } \{m_1, m_2, \dots, m_k\} \subset M$

$t.L = \text{FindLocality}(S, t.M_{\text{current}})$

$t.D = \text{FindDimensions}(k, l, t.L)$

$t.C = \text{AssignPoints}(t.D)$

$t.\text{ObjectiveFunction} = \text{EvaluateClusters}(t.C, t.D)$

if $\text{ObjectiveFunction} < \text{BestObjective}$ **then**

begin

$\text{BestObjective} = t.\text{ObjectiveFunction}$

$M_{\text{best}} = t.M_{\text{current}}$

$C_{\text{best}} = t.C$

end

end

{2. Iterative Phase}

repeat

for agent $t \in T$ **do**

begin

{find better solution with slightly change medoid}

if $rv \sim u(0,1) < 0.5$ **then**

begin

if bad medoids in $t.M_{\text{current}}$ **then**

$t.M_{\text{current}} = \text{Change } M_{\text{best}}\text{'s bad medoid with random points from } M$

else then

$t.M_{\text{current}} = \text{Change } M_{\text{best}}\text{'s one medoid with random points from } M$

end

{find better solution from totally random medoids}

else then $t.M_{\text{current}} = \text{Random set of medoids } \{m_1, m_2, \dots, m_k\} \subset M$

$t.M_{\text{current}} = \text{Random set of medoids } \{m_1, m_2, \dots, m_k\} \subset M$

$t.L = \text{FindLocality}(S, t.M)$

$t.D = \text{FindDimensions}(k, l, t.L)$

$t.C = \text{AssignPoints}(t.D)$

$t.\text{ObjectiveFunction} = \text{EvaluateClusters}(t.C, t.D)$

if $\text{ObjectiveFunction} < \text{BestObjective}$ **then**

begin

$\text{BestObjective} = t.\text{ObjectiveFunction}$

$M_{\text{best}} = t.M_{\text{current}}$

end

end

until (termination_criterion)

{3. Refinement Phase}

$L = C_{\text{best}}$

$(D_1, \dots, D_k) = \text{FindDimensions}(k, l, L)$

$(C_1, \dots, C_k) = \text{AssignPoints}(D_1, \dots, D_k)$

$N = (M_{\text{best}}, D_1, \dots, D_k)$

return(N)

end

Algorithm FindLocality(Set of points S , Set of medoids M)
begin
for each medoid $m_i \in M$ **do**
 begin
 $\delta_i = \min\{\text{Distance}(m_i, m_j, D) \mid i \neq j\}$
 $L_i = \{x \mid x \in S \text{ and } \text{Distance}(x, m_i, D) < \delta_i\}$
 end
return (L_1, \dots, L_k)

Algorithm Greedy(Set of points S , Number of medoids: k)
begin
 $M = \{m_1\}$, $\{m_1$ is a random point of $S\}$
for each $x \in S \setminus M$ **do** $\text{dist}(x) = \text{Distance}(x, m_1, D)$
for $i = 2$ to k **do**
 begin
 let $m_i \in S \setminus M$ be s. t. $\text{dist}(m_i) = \max\{\text{dist}(x) \mid x \in S \setminus M\}$
 $M = M \cup \{m_i\}$
 for each $x \in S \setminus M$ **do** $\text{dist}(x) = \min\{\text{dist}(x), \text{Distance}(x, m_i)\}$
 end
return M
end

Algorithm Distance(point x_1 , point x_2 , dimension set D_x)
begin
 $\text{CountNum} = 0$
for dimension $d \in D_x$ **do**
 begin
 if $x_1(d) = x_2(d)$
 $\text{CountNum} = \text{CountNum} + 1$
 end
return $(\text{CountNum} / |D_x|)$
end

Algorithm FindDimensions(k, l, L)
begin
 $\{X_{i,j}$ is the average distance from the points in L_i to medoid m_i , along dimension $j\}$
for each medoid i **do**
 begin
 $Y_i = \frac{\sum_{j=1}^{|D|} X_{i,j}}{|D|}$
 $D_i = \emptyset$
 $\sigma_i = \sqrt{\frac{\sum_{j=1}^{|D|} (X_{i,j} - Y_i)^2}{d-1}}$
 for each dimension j **do** $Z_{i,j} = (X_{i,j} - Y_i) / \sigma_i$
 end
 Pick the $k-l$ numbers with the least (most negative) values of $Z_{i,j}$
 subject to the constraint that there are at least 2 dimensions for each clusters
if $Z_{i,j}$ is picked **then** add dimension j to D_i
return (D_1, D_2, \dots, D_k)
end

Algorithm *AssignPoints*(D_1, D_2, \dots, D_k)
begin
for each $i \in \{1, \dots, k\}$ **do** $C_i = \emptyset$
for each $p \in P$
 Find i with the lowest value of $\text{Distance}(p, m_i, D_i)$ and add p to C_i
return (C_1, \dots, C_k)
end

Algorithm *EvaluateClusters*($C_1, \dots, C_k, D_1, \dots, D_k$)
begin
for each C_i **do**
 begin
 for each dimension $j \in D_i$ **do**
 begin
 $Y_{i,j} = \text{CentroidDistance}(C_i, d_j)$
 end
 $w_i = \frac{\sum_j Y_{i,j}}{|D_i|}$
 end
return $\left(\frac{\sum_{i=1}^k |C_i| \cdot w_i}{|P|} \right)$

Algorithm *CentroidDistance*(C_i, d_j)
 $\{d_j \text{ is a dimension in } D_i\}$
 $\text{Centroid}(C_i, d_j)$ is an average vector of $\text{OneHotEncoding}(p, d_j) \forall p \in C_i$
for each $p \in C_i$ $\text{dist}(p) = \|\text{OneHotEncoding}(p, d_j) - \text{Centroid}(C_i, d_j)\|_1 / 2$
return $\left(\frac{\sum_{p \in C_i} \text{dist}(p)}{|C_i|} \right)$