CLUSTERING AND SOFT COMPUTING

Petr Dostál

Brno University of Technology, European Polytechnic Institute, Kunovice

Abstract

Clustering has had successful applications in many fields. There are various methods of soft computing methods possible to use for this purposes. It could be fuzzy logic, neural networks and genetic algorithms. The case study of the use of soft computing clustering methods and their comparison are mentioned in the article.

1 Introduction

The soft computing plays very important roles also in clustering. The application of the soft computing clustering methods is realized on the case study. Popular notions of clusters include groups with low distances among the cluster members. The clustering could be done by fuzzy logic, networks and/or genetic algorithms. The program MATLAB® with Fuzzy Logic, Neural Network and Global Optimization Toolbox are used.

2 Theory

The fuzzy *c*-means algorithm attempts to partition a finite collection of *n* elements $X=\{x_1, x_2, \ldots, x_n\}$ into a collection of *c* fuzzy clusters with respect to some given criterion. Given a finite set of data, the algorithm returns a list of *c* cluster centres where each element and a partition matrix $W = w_{ij} \in [0,1]$, $i=1,2, \ldots, n$, $j = 1, 2, \ldots, c$, where each element w_{ij} tells the degree to which element x_i belongs to cluster c_j. The fuzzy *c*-means aims to minimize an objective function. The standard function is

$$w_k(x) = \frac{1}{\sum_j \left(\frac{d(center_{k,x})}{d(center_{j,x})}\right)^{2/(m-1)}},$$

this differs from the k-means objective function by the addition of the membership values u_{ij} and the fuzzifier *m*.

The neural network algorithm means calculation according to the formula

$$D_j = \sum_{i=1}^{n} \left(x_i(t) - w_{ji}(t)\right)^2 \rightarrow \min$$

for

$$j = 1, \ldots, N.$$

The ideal case is the situation when

$$\min_j D_j \rightarrow 0,$$

The adaptation of weighted coefficients is done according to the formula

$$w_{ji}(t+1) = w_{ji}(t) + \alpha(t)\left(x_i(t) - w_{ji}(t)\right)$$

for

$$j \in NE_C(t) \text{ and } w_{ji}(t+1) = w_{ji}(t),$$

for

$$j \notin NE_C(t).$$

The genetic algorithm obtains the calculation between an object and a centroid can be calculated by means of common Euclidean distances

$$D_E(\mathbf{x}_p, \mathbf{x}_q) = \sqrt{\sum_{l=1}^{K} (x_{pl} - x_{ql})^2} = \|\mathbf{x}_p - \mathbf{x}_q\|.$$

The aim is to find a matrix $W^* = [w_{*ij}]$ that minimizes the sum of the squares of distances in groups from their centroids (over all M centroids), i.e.

$$S(W^*) = \min_W \{S(W)\}.$$

3 Soft Computing Clustering

The application of soft computing clustering methods is realized on the cases study of 14 objects. The solved clustering is based on sorting of items according their coordinates x, y, z. See Table 1.

Coordinates			Assignment of clusters			Symbols
x	y	z	FL	NN	GA	Graph
0	16	16	2	2	2	×
34	0	0	2	2	2	×
39	26	26	2	2	2	×
35	49	49	3	3	3	♦
50	36	36	3	3	3	♦
46	48	48	3	3	3	♦
51	83	83	1	1	1	*
52	99	52	1	1	1	*
66	36	66	3	3	3	♦
81	61	81	1	1	1	*
64	95	64	1	1	1	*
85	100	85	1	1	1	*
93	98	93	1	1	1	*
100	56	100	1	1	1	*

Table 1. Case study - data

The software MATLAB and its Fuzzy Logic Toolbox is used for the software applications. The example presents the objects recorded in MS Excel format in *DC.xlsx* file. This task is solved by the program *CFL.m*. See Program 1.

```
fd=xlsread('DC.xlsx','Data');
plot3(fd(:,1),fd(:,2), fd(:,3), 'o','color','k', 'markersize',7,'LineWidth',2)
title('Data');
xlabel('x');ylabel('y');zlabel('z')
grid
 [center,U,objFcn] = fcm(fd,3);
U
objFcn
figure
plot(objFcn)
title('Fitness Function Values')
xlabel('Iteration Count')
ylabel('Fitness Function Value')
maxU = max(U);
index1 = find(U(1, :) == maxU);
index2 = find(U(2, :) == maxU);
index3 = find(U(3, :) == maxU);
figure
center
c1='x'
fd(index1,:)
c2='d'
fd(index2,:)
c3='*'
fd(index3,:)
stem3(fd(:,1),fd(:,2), fd(:,3), 'o','color','k', ...
'markersize',7)
grid
hold on
grid
stem3(center(1,1),center(1,2),center(1,3),'marker', 'x', ...
'color','g','markersize',10,'LineWidth',2)
stem3(center(2,1),center(2,2),center(2,3),'marker', ...
'd','color','r','markersize',10,'LineWidth',2)
stem3(center(3,1),center(3,2),center(3,3),'marker', ...
'*','color','b','markersize',10,'LineWidth',2)
view(30,30)
line(fd(index1, 1), fd(index1,2), fd(index1,3), ...
'linestyle','none','marker', '+','color','g');
line(fd(index2,1),fd(index2,2), fd(index2,3), ...
'linestyle','none','marker', 'd','color','r');
line(fd(index3,1),fd(index3,2), fd(index3,3), ...
'linestyle','none','marker', '*','color','b');
title('Clustering - Fuzzy Logic');
xlabel('y');ylabel('x');zlabel('z')
```

Program 1. M-file *CFL.m*

The program is started using the command *CFL* in the MATLAB program environment. The number of clusters is set up to 3. During the calculation the iteration count is displayed. When the calculation is finished the output results, the coordinates of centroids and assign of product to centroids are displayed. See Result 1.

```
center =
22.5352   12.2287   12.2509
```

```
48.8076   46.0091   49.7997
 77.6970   86.1871   81.6455
CFA = 2 2 2 3 3 3 1 1 3 1 1 1 1 1
fval: 327.1500
```

Results 1. Results of calculation

The software MATLAB and its Neural network Toolbox is used for the software applications. The example presents the objects recorded in MS Excel format in *DC.xlsx* file. This task is solved by the program *CNN.m*. See Program 2.

```
clear all;
P=(xlsread('DC','Data'))';
num=input('Zadej počet skladů:');
epochy=input('Zadej počet epoch:');
net=newc([0 1; 0 1; 0 1],num,0.1);
net.trainParam.epochs = 1;
for k=1:epochy
net = train(net,P);
w = net.IW{1};
stem3(w(:,1),w(:,2),w(:,3),'sr','MarkerFaceColor','b','MarkerSize',10)
grid on;
hold on
Q=P';
for i=1:size(Q,1)
   for j=1:(size(w,1))
      vzdalenosti(j)=sqrt((Q(i,1)-w(j,1))^2+(Q(i,2)-w(j,2))^2+(Q(i,3)-w(j,3))^2);
   end
   [min_vzdalenost(i),prirazeni(i)]=min(vzdalenosti);
end
for i=1:size(Q,1)
stem3(Q(i,1),Q(i,2),Q(i,3),'sr','MarkerFaceColor',[prirazeni(i)/num,prirazeni(i)/num,prirazeni(i)/num
],'MarkerSize',10)
xlabel('x');ylabel('y');zlabel('z');
title('Clustering - Neural Network');

end
figure(gcf)
hold off
end
celk_vzdalenost=sum(min_vzdalenost)
Q
w
prirazeni
```

Program 2. M-file *CNN.m*

The program is started using the command *CNN* in the MATLAB program environment. The number of clusters is set up to 3. During the calculation the iteration count is displayed. When the calculation is finished the output results, the coordinates of centroids and assign of product to centroids are displayed. See Result 2.

```
center =
   73.2910   86.2987   78.3818
  24.8571   13.7805   13.7805
```

```
  49.2028   42.4213   49.8212
CGA = 2 2 2 3 3 3 1 1 3 1 1 1 1 1
      fval: 329.4210
```

Results 2. Results of calculation

The software MATLAB and its Global Optimization Toolbox is used for the software applications. The example presents the objects recorded in MS Excel format in *DC.xlsx* file. This task is solved by the program *CGA.m*. See Program 3.

```
function CGA
global LOCATION;
num=input('Number of groups:');
num=3*num;
PopSize=input('Population size:');
FitnessFcn = @Group;
numberOfVariables = num;
LOCATION=(xlsread('DC','Data'))
my_plot = @(Options,state,flag) Draw(Options,state,flag,LOCATION,num);
Options = gaoptimset('PlotFcns',my_plot,'PopInitRange',[0;300],'PopulationSize',PopSize);
[x,fval] = ga(FitnessFcn,numberOfVariables,Options);
assign=zeros(1,size(LOCATION,1));
for i=1:size(LOCATION,1)
   distances=zeros(num/3,1);
   for j=1:(size(x,2)/3)
      distances(j)=sqrt((LOCATION(i,1)-x(j))^2+(LOCATION(i,2)-
x(size(x,2)/3+j))^2+(LOCATION(i,3)-x(2*size(x,2)/3+j))^2);
   end
   [min_distance,assign(i)]=min(distances);
end
assign
fval
xy=zeros(num/3,3);
for i=1:(num/3)
   xy(i,1)=x(1,i);
   xy(i,2)=x(1,num/3+i);
   xy(i,3)=x(1,2*num/3+i);
end
xy
```

Program 3. M-file *CGA.m*

The program is started using the command *CGA* in the MATLAB program environment. The number of clusters is set up to 3. During the calculation the iteration count is displayed. When the calculation is finished the output results, the coordinates of centroids and assign of product to centroids are displayed. See Result 3.

```
center =
 81.6022   59.6018   82.1812
 64.6368   95.9401   69.0854
39.5317   32.4349   33.6140
CGA = 2 2 2 3 3 3 1 1 3 1 1 1 1 1
 fval: 325.9818
```

Results 3. Results of calculation

4 Conclusion

The programs display the graphs where each object is represented by circle with symbol inside ×,♦,* that assignees objects to the clusters. The centroids of clusters are represented by symbols ×,♦,* without circle. See Figure 1. See also right columns Cluster of Table 1 (Assignment of Clusters).

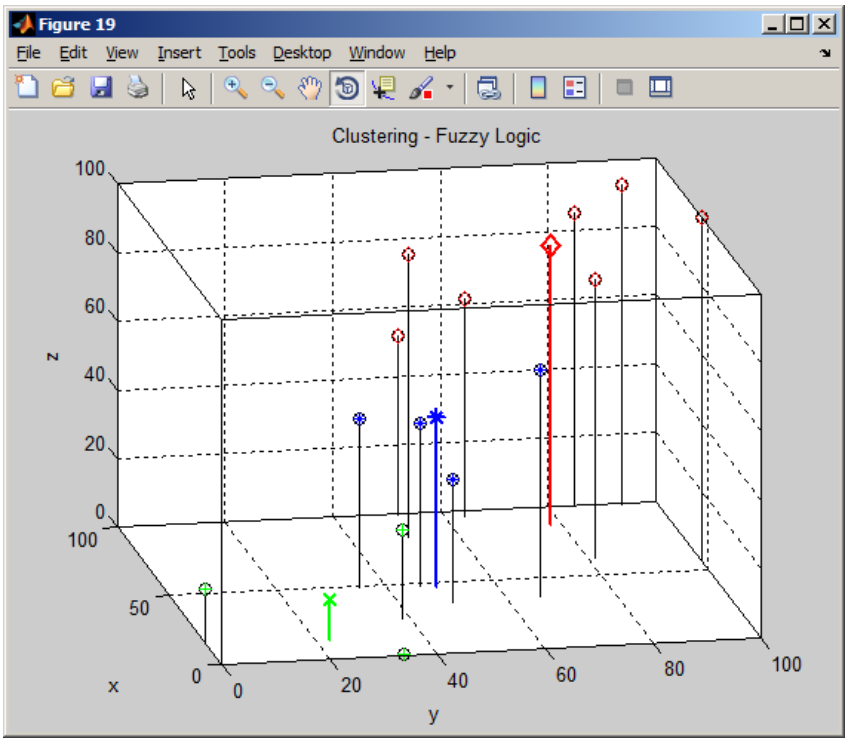

Figure 1. Three-dimensional graph – Soft Computing Clustering

The use of soft computing clustering methods (fuzzy logic, neural networks, and genetic algorithms) results in nearly the same value of fitness function, there is the same assignment of objects to the clusters, but the coordinates of centroids differs. The speediest calculation was done by fuzzy logic, then by genetic algorithms and the worst was neural network. If the number of objects will by thousands, the results could be different and it has not been tested yet.

References

[1] Bezdek, James C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. USA: Springer. 1981.

[2] Dostál, P. *Advanced Decision Making in Business and Public Services*. Czech Republic: CERM Academic Publishing House. 2011.

[3] Dostál, P. The use of soft computing for optimization in business, economics, and finance. *Meta-Heuristics Optimization Algorithms in Engineering, Business, Economics, and Finance*, USA: IGI Globe. 2012a.

[4] Dostál, P. The use of optimization methods in business and public services. *Handbook of Optimization*, USA: Springer. 2012.

[5] Dostál, P. The Use of Soft Computing in Management. *Handbook of Research on Novel Soft Computing Intelligent Algorithms: Theory and Practical Applications*, USA: IGI Globe. 2013.

[6] Vasant, P. (2003). Application of fuzzy linear programming in production planning, *Fuzzy Optimization and Decision Making*, 2 (3), 2003.

Petr Dostál
Faculty of Business and Management, Department of Informatics, Kolejní 4, 612 00 Brno
European Polytechnic Institute, Osvobození 699, 686 04 Kunovice