FUZZY CLUSTERING IN REAL ESTATE

Petr Dostál

Brno University of Technology, European Polytechnic Institute, Kunovice

Abstract

There are different methods used as supports for decision-making processes. Fuzzy logic has had successful applications in business. In this article the fuzzy clustering method is used in the field of real estates. The fundamental of fuzzy logic clustering is mentioned. The case of use is presented on real estate's clustering by means of fuzzy logic.

1 Introduction

The soft computing plays very important roles also in real estate decision making. The application of the fuzzy clustering is realized on the case of real estate grouping. Popular notions of clusters include groups with low distances among the cluster members. The fuzzy clustering could be used, not only neural networks or evolutionary algorithms. The program MATLAB® with Fuzzy Logic Toolbox is used.

2 Theory

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). In hard clustering, data is divided into distinct clusters, where each data element belongs to exactly one cluster. In fuzzy clustering (also referred to as soft clustering), data elements can belong to more than one cluster, and associated with each element is a set of membership levels. These indicate the strength of the association between that data element and a particular cluster. Fuzzy clustering is a process of assigning these membership levels, and then using them to assign data elements to one or more clusters.

One of the most widely used fuzzy clustering algorithms is the fuzzy *c*-means algorithm [3]. 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 centers 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 fuzzifier *m* determines the level of cluster fuzziness. A large *m* results in smaller memberships w_{ij} and hence, fuzzier clusters. In the limit $m = 1$, the memberships w_{ij} converge to 0 or 1, which implies a crisp partitioning. In the absence of experimentation or domain knowledge, *m* is commonly set to 2. The basic fuzzy *c*-means algorithm, given *n* data points $(x_1, x_2, \ldots , x_n)$, to be clustered, a number of *c* clusters with $(c_1, c_2, \ldots , c_n)$ and *m* the level of cluster fuzziness with.

In fuzzy clustering, every point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Thus, points on the edge of a cluster may be in the cluster to a lesser degree than points in the center of cluster. An overview and comparison of different fuzzy clustering algorithms is available.

Any point x has a set of coefficients giving the degree of being in the k-th cluster $w_k(x)$. With fuzzy c-means, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster

$$c_k = \frac{\sum_x w_k(x)x}{\sum_x w_k(x)}.$$

The degree of belonging, $w_k(x)$, is related inversely to the distance from x to the cluster center as calculated on the previous pass. It also depends on a parameter m that controls how much weight is given to the closest center.

3 Real Estate Fuzzy Clustering

The application of fuzzy clustering is realized on the cases study of real estates. The solved clustering is based on sorting of real estate's according their parameters. In other words, we have to find the real estate's with similar parameters. The variables are as follows: *Price*, *Region* and *Area*. Data are represented by 46 objects. See Table 1.

Order	*Region*	*Area*	*Price*	*Cluster*
1	21	242	600	♦
2	12	1043	1650	×
3	9	113	550	♦
4	10	929	1900	×
5	10	446	960	*
6	16	511	1300	*
7	21	366	800	*
8	10	151	360	♦
9	11	371	780	*
10	10	223	800	*
11	9	221	680	*
12	15	232	750	*
13	21	199	613	♦
14	21	214	620	♦
15	12	223	500	♦
16	19	232	515	♦
17	5	186	630	♦
18	21	187	501	♦
19	15	167	520	♦
20	10	260	900	*
21	26	167	460	♦
22	13	1906	2200	×
23	28	141	425	♦
24	5	1120	1800	×
25	26	177	450	♦
26	15	164	350	♦
27	19	186	370	♦

28	14	149	288	♦
29	17	145	300	♦
30	15	121	375	♦
31	10	465	1020	*
32	15	120	310	♦
33	15	280	520	♦
34	10	1250	1888	×
35	13	488	1160	*
36	14	394	847	*
37	10	233	575	♦
38	11	235	733	*
39	9	221	680	*
40	11	144	477	♦
41	15	164	350	♦
42	14	149	288	♦
43	16	154	434	♦
44	20	223	568	♦
45	21	203	521	♦
46	27	162	445	♦

Table 1. Real estate's data

The output will be the classification of real estate's according their characteristic to clusters. 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 *FCr.xlsx* file. This task is solved by the program *FCr.m*. See Program 1.

```
ffd=xlsread('FCr.xlsx','data');
plot3(fd(:,1),fd(:,2), fd(:,3), 'o','color','k', 'markersize',7,'LineWidth',2)
title('Data');
xlabel('Region');ylabel('Area');zlabel('Price')
grid
 [center,U,objFcn] = fcm(fd,3);
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,:)
plot3(fd(:,1),fd(:,2), fd(:,3), 'o','color','k', ...
'markersize',7)
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('Real Estate Fuzzy Clustering');
xlabel('Region');ylabel('Area');zlabel('Price'))
```

Program 1. M-file *FCr.m*

The program is started using the command *FCr* 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.

```
Iteration count = 1, obj. fcn = 6473942.582957
Iteration count = 2, obj. fcn = 4990849.902495
Iteration count = 3, obj. fcn = 4119520.114002
……
Iteration count = 36, obj. fcn = 1700291.226249
Iteration count = 37, obj. fcn = 1700291.226236
Iteration count = 38, obj. fcn = 1700291.226229

center =  1.0e+03 *
            0.0163    0.1668    0.4281
            0.0098    1.2087    1.864
           0.0126    0.3384    0.8534

c1 = ♦
ans =
   21.0000  242.0000  600.0000
    9.0000  113.0000  550.0000
   10.0000  151.0000  360.0000
   21.0000  199.0000  613.0000
   21.0000  214.0000  620.0000
   12.0000  223.0000  500.0000
   19.0000  232.0000  515.0000
    5.0000  186.0000  630.0000
   21.0000  187.0000  501.0000
   15.0000  167.0000  520.0000
   26.0000  167.0000  460.0000
   28.0000  141.0000  425.0000
   26.0000  177.0000  450.0000
   15.0000  164.0000  350.0000
   19.0000  186.0000  370.0000
   14.0000  149.0000  288.0000
   17.0000  145.0000  300.0000
   15.0000  121.0000  375.0000
   15.0000  120.0000  310.0000
   15.0000  280.0000  520.0000
```

```
  10.0000  232.9000  575.0000
  11.3333  143.6667  476.6667
  15.0000  164.0000  350.0000
  14.0000  148.6000  288.0000
  15.5000  154.4500  433.7500
  20.0000  222.8500  567.5000
  20.5000  203.3250  521.0000
  26.6667  161.5000  445.0000

c2 = ×
ans =1.0e+03 *
   0.0120    1.0430    1.6500
   0.0100    0.9290    1.9000
   0.0130    1.9060    2.2000
   0.0050    1.1200    1.8000
   0.0100    1.2495    1.8875

c3 = *
ans =1.0e+03 *
   0.0100    0.4460    0.9600
   0.0160    0.5110    1.3000
   0.0210    0.3660    0.8000
   0.0110    0.3710    0.7800
   0.0100    0.2230    0.8000
   0.0090    0.2210    0.6800
   0.0150    0.2320    0.7500
   0.0100    0.2600    0.9000
   0.0100    0.4650    1.0200
   0.0130    0.4878    1.1600
   0.0140    0.3943    0.8467
   0.0107    0.2353    0.7333
   0.0090    0.2210    0.6800
```

Results 1. Results of calculation

The program displays the graph where each real estate is represented by circle according its *Price*, *Region* and *Area*. See Figure 1.

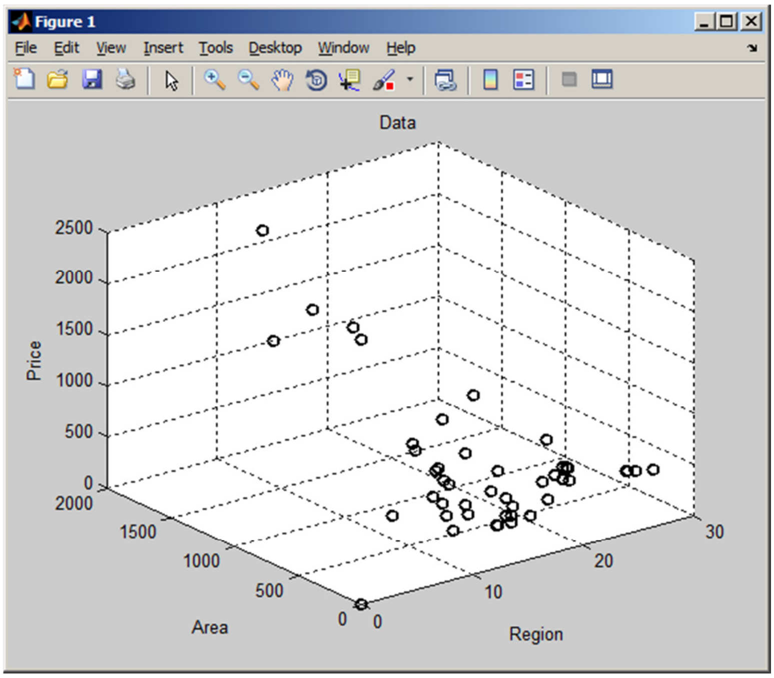

Figure 1. Three-dimensional graph – real estates

It is suitable to search the fitness function values dependent on number of iteration. The graph presents good process of iteration. See Figure 2.

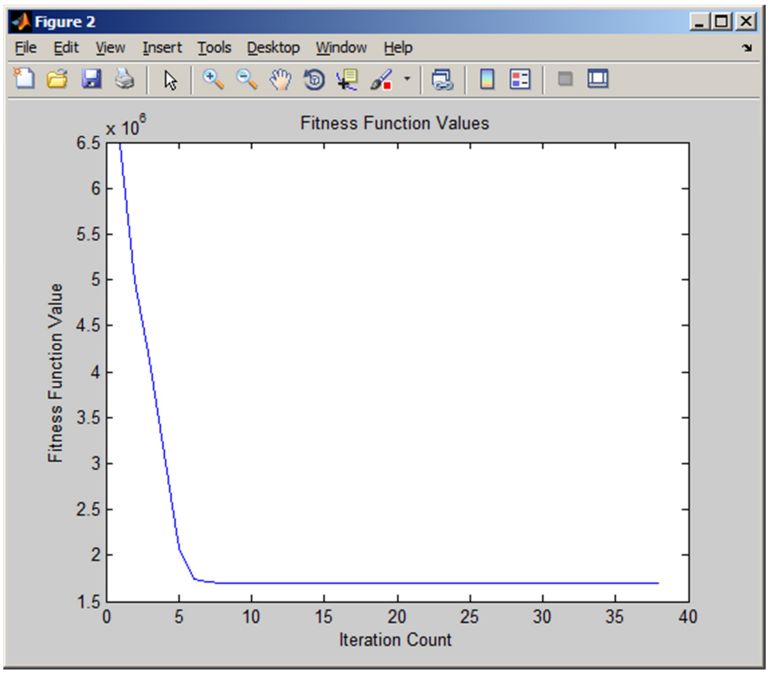

Figure 2. Fitness function values

The results are presented by coordinates of clusters and assignment of real estates to the clusters. A three-dimensional stem graph is drawn. See Figure 3. See also right column Cluster of Table 1.

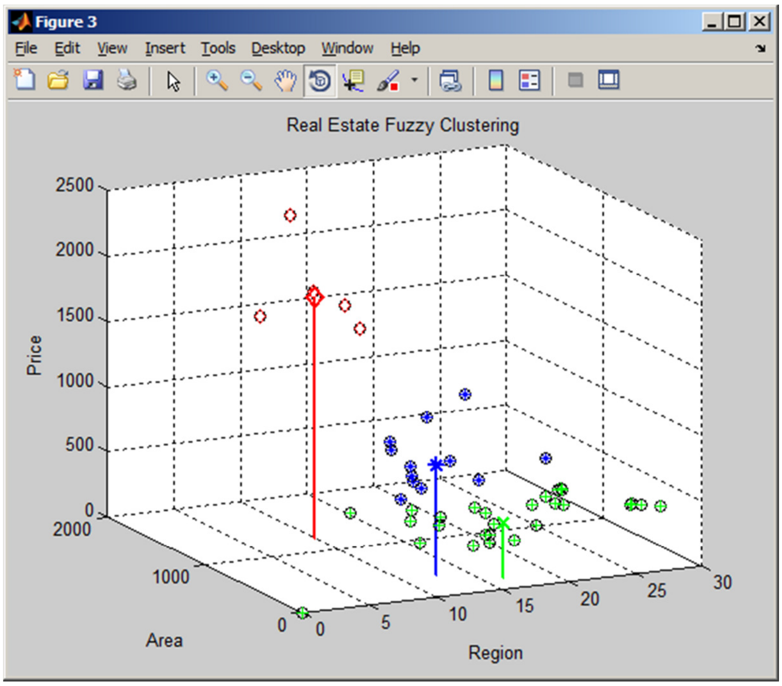

Figure 3. Graph – Real estate clustering

4 Conclusion

The results are presented by centroids of three clusters marked ♦, ×, ∗ and assignment of real estates to the clusters. The results presents the case where the cluster ♦ includes the group of real estates of middle price, good region and middle area, the cluster ∗ includes the group of real estate's of high price, in excellent region and high area, the cluster × includes the real estate's low price, bad region and small area. The fuzzy model enables to cluster the real estate's according their price, region and area. The tasks from practice lead to multi-dimensional ones, where their graphical presentation is impossible: the image of the solution is in a hyper sphere, when the variables could be price, region, area, type of real estate, number of rooms, number of floors, equipment of real estate etc. The example mentioned above is an application of the use of fuzzy logic for decision making of real estate's managers.

References

[1] Aliev, A., Fazlollahi, B., & Aliev, R. (2004). *Soft computing and its applications in business and economics*. USA: Springer.

[2] Altroc, C. (1996). *Fuzzy logic & neurofuzzy - applications in business & finance*. USA: Prentice Hall.

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

[4] Chen, S., Wang, P., & Wen, T. (2004). *Computational intelligence in economics and finance*. USA: Springer.

[5] Chen, S., Wang, P., & Wen, T. (2007). *Computational intelligence in economics and finance*. Volume II, USA: Springer.

[6] Dostál, P. (2011). *Advanced decision making in business and public services*. Czech Republic: CERM Academic Publishing House.

[7] Dostál, P. (2012a). 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.

[8] Dostál, P. (2012b). The use of optimization methods in business and public services. *Handbook of Optimization*, USA: Springer.
[9] Dostál, P. (2013a). The use of soft computing methods for forecasting in business, their applications in practice. *Nostradamus: Modern Methods of Prediction, Modeling and Analysis of Nonlinear Syst*ems, USA: Springer.
[10] Dostál, P. (2013b). The Use of Soft Computing in Management. *Handbook of Research on Novel Soft Computing Intelligent Algorithms: Theory and Practical Applications*, USA: IGI Globe.
[11] Gil-Lafuente, A.M., Gil-Lafuente, J., & Merigó-Lindahl, J.M., (2012). *Soft computing in management and Business Economics*, USA: Springer.
[12] Ribeiro, R., & Yager, R. (1999). *Soft Computing in Financial Engineering*, USA: Springer.
[13] Ruan, D., & Fedrizzi, M. (2001) *Soft Computing for Risk Evaluation and Management*, vol 76, USA: Springer.
[14] Vasant, P. (2003). Application of fuzzy linear programming in production planning, *Fuzzy Optimization and Decision Making*, 2 (3), pp. 229-241.
[15] Zadeh, L.A. (1965). Fuzzy sets. *Information and Control* 8, 338-353.
[16] Zadeh, L.A. (2012). *A Definition of Soft Computing* - adapted from L.A. Zadeh. Retrieved January 10, 2013, from http://www.soft-computing.de/def.html.

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