Dynamic Segmentation Of Behavior Patterns Based On Quantity Value Movement Using Fuzzy Subtractive Clustering Method

Iriansyah Sangadji, Yozika Arvio and Indrianto
Sekolah Tinggi Teknik PLN, Jakarta
iriansyah@sttpln.ac.id, yozika@sttpln.ac.id, indrianto@sttpln.ac.id

Abstract. to understand by analyzing the pattern of changes in value movements that can dynamically vary over a given period with relative accuracy, an equipment is required based on the utilization of technical working principles or specific analytical method. This will affect the level of validity of the output that will occur from this system. Subtractive clustering is based on the density (potential) size of data points in a space (variable). The basic concept of subtractive clustering is to determine the regions in a variable that has high potential for the surrounding points. In this paper result is segmentation of behavior pattern based on quantity value movement. It shows the number of clusters is formed and that has many members.

1. Background.
In order to understand by analyzing the pattern of changes in value movements that can dynamically vary over a given period with relative accuracy, an equipment is required based on the utilization of technical working principles or specific analytical method. This will affect the level of validity of the output that will occur from this system.

There are many techniques and methods in order to analyze the pattern of dynamically changing value movements as mentioned above. One of them is the method used based on the use of cluster to do the analysis. Clustering can also be done by utilizing various known models such as Self Organizing Mapping, Linear vector Quantitation, K Means and so on. Of the many such methods, not many are using fuzzy subtractive clustering in Clustering and classification in published research.

In this paper, it will use fuzzy subtractive clustering method in order to understand by analyzing the pattern of change in value movements of sales data in particularly. In this experiment will utilize the actual time series sales data as raw data to be simulated then analysis of simulation output results based on fuzzy subtractive clustering method.

The data used as raw data is the rough data of sales of various products within a company whose business scope is of a special characteristic.
The table 1 below is the total value of annual sales volume of the data to be simulated.

Table 1. Example times series annual sales data

| No | Year | Volume        |
|----|------|---------------|
| 1  | 2011 | Rp. 14,566,101,153 |
| 2  | 2012 | Rp. 18,222,610,140 |
| 3  | 2013 | Rp. 26,112,686,484 |
| 4  | 2014 | Rp. 28,620,582,707 |
| 5  | 2015 | Rp. 49,603,557,816 |
Segmentation can be intended as a division into different groups (heterogeneous) into homogeneous groups, where each group can be exposed to patterns of changes in varied value movements. This can be done by looking at changes in the formation of clusters according to the steps contained in the fuzzy subtractive clustering method. Number of clusters will be formed without specified number of clusters first. The algorithm with the input of a specified radius distance will determine the number of clusters to be formed.

The above will explain that the smaller the defined radius, the more clusters will be formed. On the other hand, that the larger the radius is defined, the less cluster will be formed as well. The pattern of dynamically changing value movements can be defined as cluster segmentation. Each set of resulting clusters, can contain the number of set members that can not be the same.

2. Objective
The objective of this paper is; Applying substractive clustering fuzzy method in forming cluster in time series data and Conducting segmentation process in order to get picture of behavior pattern based on movement of value of quantity.

Segmentation is a division of groups that have different characteristics, or different behaviors within a certain value. Segmentation can also be interpreted as identifying difference analysis. Segmentation can be intended as a distinct (heterogeneous) division into homogeneous groups. The purpose of segmentation is to make it easier to distinguish, meaning every value / cluster produced is easier to distinguish.

Clustering is a process of classifying data in a class or clusters so that data in a cluster has a high degree of similarity with each other but very different from the data in other clusters. Clustering is one of unsupervised learning techniques where we do not need to practice the method.

The purpose of this method is to group with number of data or objects into the cluster so that each cluster will be filled as closely as possible data. Clustering model analysis is a technique of multivariable analysis used to group objects (variables or data) so as to produce an information to assist the implementation of testing of the object and can present a hypothesis based on the relation. The principle is to maximize homogeneity in one group and to maximize the heterogeneity (inequality) between groups.

Fuzzy clustering will determine the cluster center by marking the average location for each cluster. In the initial conditions, the cluster center is still not accurate, but each data point has a membership degree for each cluster. By fixing the cluster center and the degree of membership of each data point repeatedly it will be seen that the center of the cluster will move towards the right location.

![Figure 1. Example clustering process](image)

3. Theoretical Study
in this paper, focuses on the implementation of the subtractive clustering fuzzy method. In theory that fuzzy subtractive clustering is based on the density (potential) size of data points in a space (variable). The basic concept of subtractive clustering is to determine the regions in a variable that has high potential for the surrounding points as shown in figure 1. The points with the most number of neighbors will be selected as cluster centers. The point that has been selected as the center of this
cluster will then be reduced its potential. Then the algorithm will choose another point that has the most neighbors to be the center of another cluster.

![Algorithm fuzzy subtractive clustering flowchart](image)

**Figure 2** Algorithm fuzzy subtractive clustering flowchart

This will be done repeatedly until all the points are tested. The Fuzzy Subtractive clustering algorithm is:

Matrix X which is the data to be in the cluster, the size of i x j, with i = the amount of data to be clustered and j = number of variables / attributes (criteria)

\[
\begin{bmatrix}
X_{11} & X_{12} & \ldots & X_{1j} \\
X_{21} & X_{22} & \ldots & X_{2j} \\
\vdots & \vdots & \ddots & \vdots \\
X_{i1} & X_{i2} & \ldots & X_{ij}
\end{bmatrix}
\]  

(1)

Determine:

a. \( r_j \) (the radius of each data attribute); \( j = 1, 2, \ldots, m \).
b. \( q \) (squash factor).
c. accept ratio.
d. reject ratio.
e. $X_{\text{min}}j$ (min allowed data in each data attribute); $j = 1,2, ..., m$.
f. $X_{\text{max}}j$ (max allowed data in each attribute of data); $j = 1,2, ..., m$.

Normalize

$$X_{ij} = \frac{X_{ij} - X_{\text{min}}j}{X_{\text{max}}j - X_{\text{min}}j}$$

(2)

$i = 1,2, ..., n; j = 1,2, ..., m$

Where:

$X_{ij}$ = data to the $j$-attribute.
$X_{\text{min}}j$ = minimum data on $j$-attribute.
$X_{\text{max}}j$ = maximum data on $j$-attribute.

Determine the initial potential of each data point

a. $i = 1$; $i$ is data to.
b. Do it up to $i = n$,
   i. $T_{j} = X_{ij}$ ($T_{j}$ is the $i$th $j$th attribute data); $j = 1,2, ..., m$
   ii. Count:

$$\text{Dist}_{ij} = \left( \frac{T_{j} - X_{\text{min}}j}{r_{j}} \right)$$

(3)

iii. Calculate initial potential:

$$D_{i} = \sum_{j=1}^{m} e^{-\left( \text{Dist}_{ij} \right)}$$

if $m = 1$, Then

$$D_{i} = \sum_{j=1}^{m} e^{-\left( \text{Dist}_{ij} \right)}$$

if $m > 1$, Then

$$D_{i} = \sum_{j=1}^{m} e^{-\left( \text{Dist}_{ij} \right)}$$

(4)

To calculate the degree of membership

$$\mu_{k_{i}} = e^{-\sum_{j=1}^{m} \frac{\left( X_{ij} - C_{ij} \right)^{2}}{2\sigma_{j}^{2}}}$$

(5)

4. Data

The data variables used in this research are:

a) Variable 1: data of quantity
b) Variable 2: net value data
c) Variable 3: data cogs
Process requirements here are variables and parameters used to process data in general. The specifications of software and hardware that the authors use in this study are Variables:

\[ X_1 = \text{Quantity} \]
\[ X_2 = \text{Net Value} \]
\[ X_3 = \text{COGS} \]

The parameter value used is the result of the experiment that has been done so that obtained the maximum parameter value.

a. \( r \) (radius) = 0.001
b. \( q \) (squash factor) = 0.2
c. accept ratio = 0.8
d. reject ratio = 0.2
e. \( X_{1\text{Min}} = 1, X_{1\text{Max}} = 90 \)
\[ X_{2\text{Min}} = 46162800, X_{2\text{Max}} = 1140000000 \]
\[ X_{3\text{Min}} = 63649002, X_{3\text{Max}} = 1014124250 \]

5. Result and Discussion
In the raw data groupings of a year data, the radius used is 0.8 resulting in 1 cluster, 0001 yielding 2 clusters and 0.0001 resulting in 3 clusters.

![Figure 3. Example of gui’s clusters](image)

Gui’s clusters has been designed using matlab as shown Figure 3 above. A year data testing is done by dividing into four quarters result data. 3-dimensional correlation of data \((X_1, X_2, X_3)\) can be seen also entirely in figure 3a,b,c.

| Cluster Center | Number of Cluster Members | Total data | Quarters       | Clusters |
|----------------|---------------------------|------------|----------------|----------|
| 1              | 7589                      | 12561      | First Quarter  | 2        |
| 2              | 2663                      | 12561      |                |          |
| 1              | 7589                      | 12561      | Second Quarter | 2        |
| 2              | 2663                      | 12561      |                |          |
| 1              | 7589                      | 12561      | Thirth Quarter | 3        |
| 2              | 2663                      | 12561      |                |          |
Table 2 shows that there is a significant change in the behavior of the data increase pattern in the third and fourth quarters.

Table 3. Net value data

| Cluster Center | Cluster Members | Number of Cluster members | Quarters   | Clusters |
|----------------|----------------|--------------------------|------------|----------|
| 78             | 217,646        | 2                        | First Quarter | 2        |
| 148            | 92,373         | 2                        | Second Quarter | 2        |
| 32             | 1721           | 1                        |            |          |
| 77200          | 4279           | 1                        |            |          |
| 196480         | 192            | 1                        |            |          |
| 1480410        | 46,49,68,110,138,271,274,299,302,305,320,383,417,432,448,465,472,481,498,530,566,567,759,761,796,811,855,888,889,895,911,982,983,984,1031,1093,1108,1126 | 38 | Thirth Quarter | 3 |
| 122268         | 573            | 1                        |            |          |
| 67191          | 424,547        | 2                        |            |          |
| 263464         | 45             | 1                        | Fourth Quarter | 3        |
| 52800          | 1385           | 1                        |            |          |
| 25400          | 5400,6317,6399,8451,8533,10630,10704 | 7 |            |          |
| 116008         | 11856,11863,12184,12190,12202,12208,12466,12524 | 8 | Total A Year | 3 |
| 30200          | 5636,5881      | 2                        |            |          |

Table 3 showed that there was a significant increase in behavior change in the third quarter. Although there is also the same thing in the fourth quarter, but not as high as third quarter.

Table 4. Cogs data

| Cluster Center | Cluster Members | Number of Cluster members | Quarters   | Clusters |
|----------------|----------------|--------------------------|------------|----------|
| 546840         | 217            | 1                        | First Quarter | 2        |
| 864672         | 92             | 1                        | Second Quarter | 2        |
| 219247         | 1721           | 1                        |            |          |
| 162779         | 4279           | 1                        |            |          |
| 106965         | 192,731        | 2                        |            |          |
| 759614         | 35,46,49,68,110,138,182,271,274,299,302,305,320,383,417,432,448,465,472,481,498,530,566,567,759,761,796,811,855,888,889,911,982,983,984,986,1031,1093,1108,1126,1135,1136,1137 | 95 | Thirth Quarter | 3 |
Table 4 the same thing as table 3, shows that there is a significant jump in behavior change in the data pattern in the third quarter. Although there is also the same thing in the fourth quarter, but not as high as third quarter.

6. Conclusion

The data grouped by the fuzzy subtractive clustering method begins with the input data to be clustered, which begins by setting the value of cluster parameters consisting of: the radius of each attribute data (r), squash factor (q), Accept ratio, reject ratio, Xmin (minimum allowed data), Xmax (minimum allowed data), data normalization, determine the initial potential of each data point, locate the highest potential point, determine the cluster center locate the new cluster center candidate and calculate the sigma cluster value. It can produce the number of clusters according to the data being implemented.

There was a significant dynamic segmentation in the third quarter where the value of the changes was quite high.

References
[1] A. Taupik Rahman, Nasrun Hariyanto, & Sabat Anwari. (2014). Peramalan Beban Puncak Jangka Pendek Khusus Hari Libur Nasional Berbasis Algoritma Fuzzy Subtractive Clustering, Studi Kasus di Jawa – Bali. Jurnal Reka Elkomika. 2337-439X April 2014.
[2] Away, A. (2010). Matlab Programming. Bandung : Informatika.
[3] Gita Febrina Wulandari. (2014). Segmentasi Pelanggan Menggunakan Algoritma K-Means Untuk Customer Relationship Management (CRM) Pada Hijab Miulan.
[4] Helmy Yulianto Hadi, R. Rizal Isnanto, & Budi Setiyono. (2006). Pembagian Kelas Kuliah Mahasiswa Menggunakan Algoritma Pengklasteran Fuzzy. Makalah Seminar Universitas Diponegoro.
[5] Heru Dibyo Laksono, & Muhammad Hafis. (2013). Aplikasi Fuzzy Clustering Dengan Menggunakan Algoritma.
[6] Subtractive Clustering Untuk Perkiraan Kebutuhan Energi Listrik Jangka Panjang Di Provinsi Sumatra Barat Dari Tahun 2012-2021. Jurnal Teknologi Informasi & Pendidikan. 2086 – 4981.
[7] I Gede Oka Artawan. (2013). Penentuan Lokasi SMP Baru di Kabupaten Klungkung dengan Algoritma Fuzzy Subtractive Clustering. Jurnal Matematika. 1693-1394.
[8] Kotler, Philip dan Keller, Kevin L. (2009), Manajemen Pemasaran edisi ke-13 jilid 1, Penerbit Erlangga, Jakarta.
[9] Kusumadewi, Sri. (2002). Analisis dan Desain Sistem Fuzzy Menggunakan Toolbox Matlab. Yogyakarta: Graha Ilmu.
[10] Kusumadewi, Sri, Hari Purnomo. (2004). Aplikasi Logika Fuzzy Untuk Pendukung keputusan. Graha Ilmu, Yogyakarta.