

Application Kohonen Network and Fuzzy C Means for Clustering Airports Based on Frequency of Flight

Dinita Rahmalia^{*1}, Teguh Herlambang²

¹Universitas Islam Darul Ulum Lamongan, ²Universitas Nahdlatul Ulama Surabaya
dinitarahmalia@gmail.com^{*1}, teguh@unusa.ac.id²

Abstract

In Indonesia, the demands of air transportation for reaching destination increase rapidly. Based on the flight schedule in airports spreading in Indonesia, the airports have different flight demand rate so that it requires clustering. This research will use two methods for clustering: kohonen network and Fuzzy C Means (FCM). Kohonen network is the type neural network which uses unsupervised training. Kohonen network uses weight vectors for training while FCM uses degree of membership. Both kohonen network and FCM, inputs are represented by the number of departure and arrival of airline in one day. For kohonen network, we update weight matrices so that minimizing the sum of optimum euclidean distance. For FCM, we update degrees of membership so that minimizing the objective function value. From the simulations, we can cluster the airports based on the number of departure and arrival of airline.

Keywords: Clustering, Neural Network, Kohonen Network, Fuzzy C Means

1. Introduction

Indonesia is the one of largest country in the world so that the demands of air transportation for reaching destination increase rapidly. Based on the flight schedule in airports spreading in Indonesia, the airports have different flight demand rate. Due to the difference of flight demand on flight among airports, the management should make clustering in some airports. Clustering will make flight schedule like the number of departure and arrival airline more proper.

In previous research, K means method has been applied in clustering problem [1]. This research will use two methods for clustering : kohonen network and Fuzzy C Means (FCM). Kohonen network is application of Neural Network (NN) and Fuzzy C Means (FCM) is application of fuzzy set.

Neural Network (NN) was introduced by McCulloch and Pitts in 1943. NN work resembles to human neuron system. There are some application of NN : prediction [2][3], clustering [4][5], and classifying [6]. Type of NN used in clustering process is kohonen network. Clustering is the process of grouping a set of data objects into multiple groups or clusters.

Kohonen network is the type neural network which uses unsupervised training. In unsupervised training, there are weight vectors. The weight vector for an output unit in a clustering net serves as a representative or exemplar vector for the input patterns which the net has placed on that cluster. During training, the net determines the output unit that is the best match for the current input vector. The weight vector for the winner is then adjusted in accordance with the net's learning algorithm [7].

Fuzzy set was introduced by L.A. Zadeh in 1965. Fuzzy set differs with crisp set in which fuzzy set has membership function [8]. There are many applications in fuzzy system: optimization, control, estimation [9]. The application of fuzzy set can be used in clustering problem using Fuzzy C Means (FCM).

Fuzzy C Means (FCM) is the data clustering method where each data belongs to a cluster to some degree that is specified by degree of membership. This method was introduced by Bezdek in 1981 [10].

Both kohonen network and FCM can cluster some airports based on the number of departure and arrival of airline. For kohonen network, we update weight matrices so that minimizing the sum of optimum euclidean distance. For FCM, we update degrees of membership so that minimizing the objective function value.

From the simulations, we can cluster the airports based on the number of departure and arrival of airline with various size of cluster.

2. Research Method

2.1 Data Clustering

Clustering is the process of grouping a set of data objects into multiple groups or clusters so that objects within a cluster have high similarity, but are very dissimilar to objects in other clusters. Dissimilarities and similarities are assessed based on the attribute values describing the objects and often involve distance measures [11].

Clustering as a datamining tool has its roots in many application areas. The purpose of clustering is to identify natural groupings of data from a large data set to produce a concise representation of a system behavior.

2.2 Supervised Training and Unsupervised Training in Neural Network

Based on weight modification, there are two types of training: supervised and unsupervised training. In supervised training, there is a pair of input and target used to train net until optimal weights are obtained. For each training, inputs are given to the net. The net will process the output. Error between output and target will be propagated for updating weights.

In unsupervised training, there are weight vectors. The weight vector for an output unit in a clustering net serves as a representative or exemplar vector for the input patterns which the net has placed on that cluster. During training, the net determines the output unit that is the best match for the current input vector. The weight vector for the winner is then adjusted in accordance with the net's learning algorithm [7].

2.3 Kohonen Network

The self organizing neural networks are called topology-preserving maps, assume a topological structure among the cluster units. There are m cluster units, the input signals are n tuples, like on Figure 1.

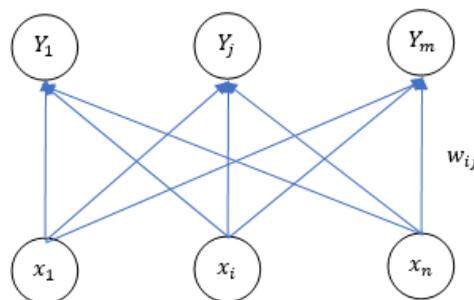


Figure 1. Kohonen Network

The weight vector for a cluster unit serves as an exemplar of the input patterns associated with that cluster. During the self organizing process, the cluster unit whose weight vector matches the input pattern most closely (the square of the minimum euclidean distance) is chosen as the winner. The winning unit and its neighboring units update their weights.

Kohonen Algorithm can be explained as follows [7]:

1. Suppose D is the number of data, N is the number of clusters. Initialize weights w_{ij} , set topological neighborhood parameters, set learning rate parameters α .

2. For $k = 1, 2, \dots, D$, do steps 2.1 until 2.3

2.1. For $j = 1, 2, \dots, N$, calculate euclidean distance $dist(j)$ on Equation 1.

$$dist(j) = \sum_i (w_{ij} - x_i)^2 \quad (1)$$

2.2. Find index J such as $dist(j)$ is a minimum.

2.3. For all units j within a specified neighborhood of J and for all i on Equation 2.

$$w_{ij}(new) = w_{ij}(old) + \alpha(x_i - w_{ij}(old)) \quad (2)$$

3. Calculate the sum of optimum euclidean distance on Equation 3.

$$\sum_{k=1}^D \|x_k - w^*\|^2 \quad (3)$$

4. Update learning rate on Equation 4.

$$\alpha(t+1) = \theta\alpha(t) \quad (4)$$

5. Reduce radius of topological neighborhood at specified times

6. Test stopping condition

2.4 Fuzzy C Means

Fuzzy C Means (FCM) is the data clustering method where each data belongs to a cluster to some degree that is specified by degree of membership. This method was introduced by Bezdek in 1981 [5]. It provides how to cluster data that populate some dimensional space into a specific number of different clusters.

FCM is the clustering method that allows each data to belong to multiple clusters with varying degrees of membership.

FCM is based on the minimization of the following objective function on Equation 5.

$$J_m = \sum_{i=1}^D \sum_{j=1}^N \mu_{ij}^m \|x_i - c_j\|^2 \quad (5)$$

Where:

D : is the number of data

N : is the number of clusters

m : is fuzzy partition matrix exponent for controlling the degree of fuzzy overlap, with $m > 1$

x_i : is the i -th data

c_j : is the center of the j -th cluster

μ_{ij} : is the degree of membership of x_i in the j -th cluster

For every given data x_i , $i = 1, 2, \dots, D$ the sum of the membership values for all clusters is one as in Equation 6.

$$\mu = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1N} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{D1} & \mu_{D2} & \cdots & \mu_{DN} \end{bmatrix} \quad (6)$$

$$\mu_{ij} \sim U(0,1), \quad i = 1, 2, \dots, D, \quad j = 1, 2, \dots, N \quad (7)$$

$$\sum_{j=1}^N \mu_{ij} = 1, \quad i = 1, 2, \dots, D \quad (8)$$

The algorithm of FCM is as follows [9] :

1. Randomly, initialize the cluster membership values μ_{ij} with constrains Equation 7 and Equation 8

2. Calculate the cluster centers on Equation 9 with $m = 2$.

$$c_j = \frac{\sum_{i=1}^D \mu_{ij}^m x_i}{\sum_{i=1}^D \mu_{ij}^m} \quad j = 1, 2, \dots, N \quad (9)$$

3. Update μ_{ij} based on Equation 10.

$$\mu_{ij} = \frac{1}{\sum_{k=1}^N \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad i = 1, 2, \dots, D \quad j = 1, 2, \dots, N \quad (10)$$

4. Calculate the objective function J_m in Equation 5.
5. Repeat step 2 – 4 until stopping condition is satisfied.

3. Results and Discussion

This section explains about the data used and two methods: kohonen network and FCM in clustering the airports.

3.1 The Data of Airport

Data used in simulations are airport data spreading in Indonesia, Kuala Lumpur, Singapore, and Bangkok. Data are taken on February 3, 2012 [12]. From each airport, we calculate the number of departure and arrival of Boeng B738 airline on the date based on flight schedule Table 1.

Table 1. Flight Schedule of Boeing B738

Flight Number	Departure Airport	Arrival Airport	Departure Time	Arrival Time
1	CGK	UPG	00.45	03.05
2	CGK	AMQ	03.50	05.35
3	CGK	MES	05.30	07.30
⋮	⋮	⋮	⋮	⋮
213	DPS	CGK	21.35	23.20
214	SUB	CGK	21.55	23.15

The number of departure and arrival of airline will affect on airport clustering. Table 2 shows airports observed.

Table 2. Code and Location of Airports

Code	Location	Code	Location	Code	Location	Code	Location
BTJ	Banda Aceh	CGK	Cengkareng	BDJ	Banjarmasin	KOE	Kupang
MES	Medan	BDO	Bandung	BPN	Balikpapan	TTE	Ternate
PKU	Pekanbaru	SRG	Semarang	MDC	Manado	AMQ	Ambon
PDG	Padang	SOC	Solo	GTO	Gorontalo	TIM	Timika
DJB	Jambi	JOG	Yogyakarta	PLW	Palu	BIK	Biak
PLM	Palembang	SUB	Surabaya	KDI	Kendari	DJJ	Jayapura
TKG	Bandar Lampung	MLG	Malang	UPG	Makasar	KUL	Kuala Lumpur
PGK	Pangkal Pinang	PNK	Pontianak	DPS	Denpasar	SIN	Singapore
BTH	Batam	PKY	Palangkaraya	LOP	Lombok	BKK	Bangkok

3.2 Simulation of Kohonen Network

In this simulation, we will apply clustering by 4, 5, 6, and 7 clusters. Parameters used in kohonen network are:

1. The number of airport data D : 36
2. The number of clusters N : 4, 5, 6, 7
3. Learning rate parameter α : 0.9
4. Maximum epoch : 100

The form of vector used in kohonen network can be represented in Equation 11 where d_i is the number of departure of airline at i -th hour and a_i is the number of arrival of airline at i -th hour.

$$x = \begin{bmatrix} d_1 & \dots & d_i & \dots & d_{24} \\ a_1 & \dots & a_i & \dots & a_{24} \end{bmatrix} \quad (11)$$

The clustering result can be seen in Table 3, Table 4, Table 5, and Table 6. Table 3 represents the clustering result by 4 clusters, Table 4 represents the clustering result by 5 clusters, Table 5 represents the clustering result by 6 clusters, and Table 6 represents the clustering result by 7 clusters.

The sum of optimum euclidean distance can be seen in Figure 2 (a) – (d). From the simulation, we can see that the weight matrices are updated so that the sum of optimum euclidean distance converges along epoch time. Figure 2(a) represents the sum of optimum euclidean distance by 4 clusters is 1.352, Figure 2(b) represents the sum of optimum euclidean distance by 5 clusters is 1.107, Figure 2(c) represents the sum of optimum euclidean distance by 6 clusters is 0.981, and Figure 2(d) represents the sum of optimum euclidean distance by 7 clusters is 0.845.

Table 3. Clustering Result by 4

Cluster	Airport Code
1	SUB, UPG
2	CGK
3	BTJ, PDG, DJB, TKG, PGK, BDO, SOC, MLG, PNK, PKY, BDJ, MDC, GTO, PLW, KDI, DPS, LOP, KOE, TTE, AMQ, TIM, BIK, DJJ, KUL, BKK
4	MES, PKU, PLM, BTH, SRG, JOG, BPN, SIN

Table 4. Clustering Result by 5

Cluster	Airport Code
1	SUB, DPS
2	BTJ, PDG, DJB, TKG, PGK, BDO, SOC, MLG, PNK, PKY, BDJ, MDC, GTO, PLW, KDI, LOP, KOE, TTE, AMQ, TIM, BIK, DJJ, KUL, BKK
3	CGK
4	MES, PKU, PLM, BTH, SRG, JOG, BPN, SIN
5	UPG

Table 5. Clustering Result by 6

Cluster	Airport Code
1	SUB, DPS
2	CGK
3	UPG
4	BTH, SOC, JOG
5	MES, PKU, PLM, SRG, BPN, SIN
6	BTJ, PDG, DJB, TKG, PGK, BDO, MLG, PNK, PKY, BDJ, MDC, GTO, PLW, KDI, LOP, KOE, TTE, AMQ, TIM, BIK, DJJ, KUL, BKK

Table 6. Clustering Result by 7

Cluster	Airport Code
1	CGK
2	BTJ, PDG, DJB, TKG, PGK, BDO, MLG, PNK, PKY, BDJ, MDC, GTO, PLW, KDI, LOP, KOE, TTE, AMQ, TIM, BIK, DJJ, KUL, BKK
3	UPG
4	BTH, BPN, SIN
5	MES, PKU, PLM, SRG
6	SOC, JOG
7	SUB, DPS

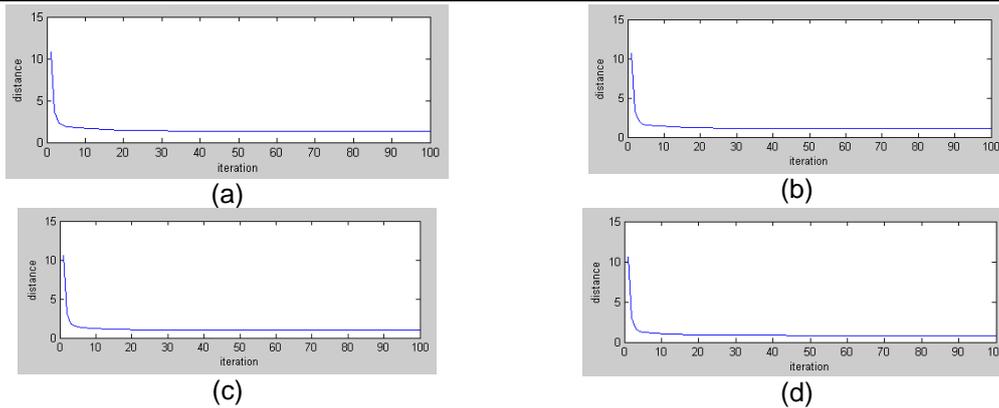


Figure 2. Optimization Process of Kohonen Network on Euclidean Distance until 100 Epoches. (a) By 4 Clusters (b) By 5 Clusters (c) By 6 Clusters (d) By 7 Clusters

3.3 Simulation of FCM

In this simulation, we will apply clustering by 4, 5, 6, and 7 clusters. Parameters used in FCM are:

1. The number of airport data D : 36
2. The number of clusters N : 4, 5, 6, 7

The form of data used in FCM can be represented in two dimensional data $x = (d, a)$ where d is the number of departure of airline on the date and a is the number of arrival of airline on the date.

The clustering result can be seen in Table 7, Table 8, Table 9, and Table 10. Table 7 represents the clustering result by 4 clusters, Table 8 represents the clustering result by 5 clusters, Table 9 represents the clustering result by 6 clusters, and Table 10 represents the clustering result by 7 clusters.

Table 7. Clustering Result by 4

Cluster	Airport Code
1	BTJ, PDG, DJB, TKG, PGK, BDO, MLG, PNK, PKY, BDJ, MDC, GTO, PLW, KDI, LOP, KOE, TTE, AMQ, TIM, BIK, DJJ, KUL, BKK
2	MES, PLM, SUB, UPG, DPS
3	CGK
4	PKU, BTH, SRG, SOC, JOG, BPN, SIN

Table 8. Clustering Result by 5

Cluster	Airport Code
1	BTJ, PDG, DJB, TKG, PGK, BDO, MLG, PNK, PKY, BDJ, MDC, GTO, PLW, KDI, LOP, KOE, TTE, AMQ, TIM, BIK, DJJ, KUL, BKK
2	MES, PLM, UPG, DPS
3	PKU, BTH, SRG, SOC, JOG, BPN, SIN
4	CGK
5	SUB

Table 9. Clustering Result by 6

Cluster	Airport Code
1	PKU, BTH, SOC
2	SUB
3	CGK
4	BTJ, PDG, DJB, TKG, PGK, BDO, MLG, PNK, PKY, BDJ, MDC, GTO, PLW, KDI, LOP, KOE, TTE, AMQ, TIM, BIK, DJJ, KUL, BKK
5	SRG, JOG, BPN, SIN
6	MES, PLM, UPG, DPS

Table 10. Clustering Result by 7

Cluster	Airport Code
1	CGK
2	BTJ, PDG, TKG, PKY, BDJ, MDC, GTO, PLW, LOP, KOE, AMQ, KUL
3	SUB
4	MES, PLM, DPS
5	UPG
6	PKU, BTH, SRG, SOC, JOG, BPN, SIN
7	DJB, PGK, BDO, MLG, PNK, KDI, TTE, TIM, BIK, DJJ, BKK

The objective function value can be seen in Figure 3 (a) – (d). From the simulation, we can see that the degrees of membership are updated so that the objective function value converges along epoch time. Figure 3(a) represents the objective function value by 4 clusters is 66.881, Figure 3(b) represents the objective function value by 5 clusters is 32.939, Figure 3(c) represents the objective function value by 6 clusters is 24.050, and Figure 3(d) represents the objective function value by 7 clusters is 11.873.

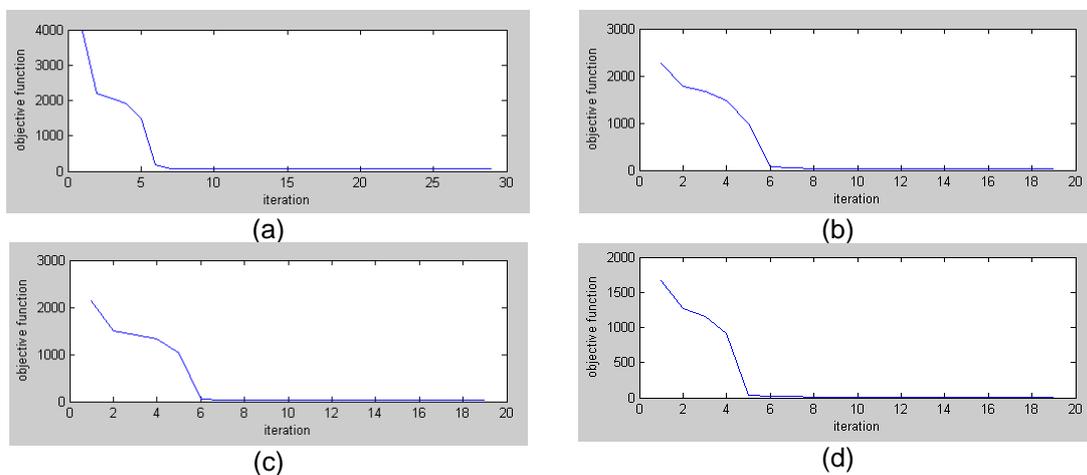


Figure 3. Optimization Process of Fuzzy C Means on Objective Function (a) By 4 clusters (b) By 5 clusters (c) By 6 clusters (d) By 7 clusters

4. Conclusion

Both kohonen network and FCM can cluster some airports based on the number of departure and arrival of airline. For kohonen network, we update weight matrices so that minimizing the sum of optimum euclidean distance. For FCM, we update degrees of membership so that minimizing the objective function value. From the simulations, we can cluster the airports based on the number of departure and arrival of airline.

References

- [1] M. Zaki, W. M. Jr, and W. Meira, *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press, 2014.
- [2] T. Rahmalia, D and Herlambang, "Prediksi Cuaca Menggunakan Algoritma Particle Swarm Optimization-Neural Network (PSONN)," in *Seminar Nasional Matematika dan Aplikasinya (SNMA)*, pp. 41–48, 2017.
- [3] J. . Siang, "Jaringan Syaraf Tiruan dan Pemrogramannya menggunakan Matlab," ANDI, 2009.
- [4] T. Kohonen, "Essentials Of The Self Organizing Map," *Neural Networks*, Vol. 37, Pp. 52–65, 2013.
- [5] T. Kohonen, "An Introduction to Neural Computing," *Neural Networks*, Vol. 1, No. 1, Pp. 3–16, 1988.
- [6] S. Ghorpade, J. Ghorpade, and S. Mantri, "Pattern Recognition Using Neural Networks,"

-
- International Journal Of Computer Science And Information Technology, Vol. 2, No. 6, Pp. 92–98, 2010.
- [7] L. Faucett, "*Fundamentals of Neural Networks*," Prentice-hall, 1994.
- [8] L. . Zadeh, "*Fuzzy Set*," Information And Control, Vol. 8, Pp. 338–353, 1965.
- [9] H. J. Zimmermann, "*Fuzzy Control*," in *Fuzzy Set Theory—and Its Applications*, Dordrecht: Springer Science+Business Media, 2001.
- [10] J.C. Bezdec, "*Pattern Recognition with Fuzzy Objective Function Algorithms*," Plenum Press, 1981.
- [11] J. Han, M. Kamber, and J. Pei, "*Data Mining Concept and Techniques*," Elsevier. Ltd, 2012.
- [12] N. K. Rahmalia. D, Hadianti. R, "*Optimisasi Crew Pairing dengan Memodifikasi Jadwal Penerbangan*," ITB, 2013.