

VISUALIZATION MODULE OF DENSITY-BASED CLUSTERING FOR HOTSPOT DISTRIBUTION IN INDONESIA USING MAPSERVER

Yenni Puspitasari*, Imas Sukaesih Sitanggang, and Rina Trisminingsih

Department of Computer Science, Bogor Agricultural University, Bogor, 16680, Indonesia

*Corresponding author: yennipsptsr@gmail.com

ABSTRACT

A web-based Geographic Information System (GIS) has been built by previous researchers to visualize hotspots data in Indonesia. That GIS still has not contained a hotspot analysis module. Data mining method can be used to analyze hotspot data. This research aims to develop and to integrate a clustering module of hotspot in GIS which has been developed in the previous research. The clustering module for grouping hotspot data was built using the DBSCAN algorithm with PHP programming language. Clustering hotspot data was performed based on year, month, and province. Clustering parameters are epsilon and minimum points (MinPts). Epsilon value that used ranged from 0.01 to 0.1 while MinPts ranges from 1 to 6. The clustering results are shown in form of table which consists of the attribute Province, Regency, Latitude, Longitude and Cluster. Cluster column is the final result of clustering using algorithm DBSCAN. The attribute cluster represents clusters are visualized using the map of Indonesia that was built using MapServer. Visualization can help parties involved in making effective and efficient decisions to prevent forest fires.

Key words: clustering, DBSCAN algorithm, geographic information system, hotspot

INTRODUCTION

Forest fire is a natural disaster that happened in Indonesia frequently. Forest fire which happened in Indonesia is caused by various processes of degradation, deforestation and climate El Nino phenomena. That various processes are changing forest area in Indonesia from ecosystem-resistant become ecosystem that vulnerable to the forest fire. Indonesia's forest fire is not only detrimental for Indonesia, but also detrimental for neighbor countries, such as Singapore and Malaysia. In 1997 – 1998, forest fires were happened in 23 of 27 provinces in Indonesia. In July 2003, there were 1 210 hotspots, and around 98% happened in Riau Province. Therefore, the prevention of forest fire in Indonesia must to be done in order to control and minimize the impacts (Purba *et al.* 2014).

Developing Geographical Information System (GIS) is one of activities to prevent forest fire. The system can be used to know the distribution of grouping hotspot data. Hotspot data become one of indicators forest fire occurrences. That data are used because a hotspot becomes a measure of the temperature of a location that has relatively higher temperature than the ambient temperature.

The previous research has built a module in GIS for searching hotspot distribution in Indonesia (Mulya 2014). That GIS which has been built by previous researcher that show hotspot distribution in Indonesia based on location and time. The summary of hotspot is visualized in form of map, table and graph. The map has integrated with GoogleMap. Map visualization is also built with GeoExplorer that provided by OpenGeo

Suite. However, this GIS still has not contained a hotspot analysis module using data mining techniques.

In this research we built a clustering module for historical data hotspot in Indonesia in GIS using Density-Based clustering (DBSCAN) algorithm. This module uses client-server architecture that is adapted from the work (Amri 2015). This module as server web is integrated with the GIS than has been built by previous research (Mulya 2014). Historical hotspot data which used are hotspot data forest fire in Indonesia from 2002 until 2012.

RESULT AND DISCUSSION

Hotspot data

Hotspot data that used in this research are hotspot data from 2002 until 2012 which were obtained from (FIRMS NASA) and Ministry of Forestry, Indonesia. That data has conducted pre-process by previous research (Mulya 2014). Pre-process includes attributes selection which are relevant to the research.

Clustering module using DBSCAN algorithm

Generally, functions that are added for completing functional requirement in this GIS are visualize the clustering hotspot result that processed with DBSCAN algorithm and visualize the clustering hotspot result in map of Indonesia in the visualization module MapServer. In detail, the DBSCAN algorithm is shown in the following pseudocode (Ester *et al.* 1996).

```

1  DBSCAN(D, eps, MinPts) C = 0
2  for each unvisited point P in dataset D mark P as visited
3  NeighborPts = regionQuery(P, eps)
4  if sizeof(NeighborPts) < MinPts mark P as NOISE else C = next cluster
5      expandCluster(P, NeighborPts, C, eps, MinPts)
6      expandCluster(P, NeighborPts, C, eps, MinPts)
7      add P to cluster C
8      for each point P' in NeighborPts
9          if P' is not visited
10             mark P' as visited
11             NeighborPts' = regionQuery(P', eps)
12             if sizeof(NeighborPts') >= MinPts
13                 NeighborPts = NeighborPts joined with NeighborPts'
14             if P' is not yet member of any cluster
15                 add P' to cluster C regionQuery(P, eps)
16  return all points within P's eps-neighborhood (including P)

```

Algorithm 1. Pseudocode of DBSCAN algorithm

DBSCAN is based on two main concepts: density reachability and density connectability. These both concepts depend on two input parameters of the DBSCAN clustering: the size of epsilon neighborhood (ϵ) and the minimum points in a cluster (m). The key idea of the DBSCAN algorithm is that, for each point of a cluster, the neighborhood of a given radius ϵ has to contain at least a minimum number of points m , that is, the density in the neighborhood has to exceed some predefined threshold (Kantardzic 2011). Figure 1 shows the illustration of density reachability concept. Figure 2 shows the illustration of density connectability concept.

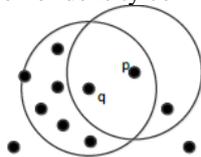


Figure 1. The illustration of density reachability concept

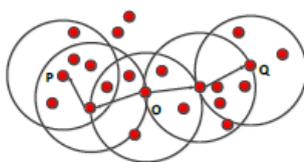


Figure 2. The illustration of density connectability concept

The clustering module for hotspots data was built using PHP programming language with the framework CodeIgniter. The implementation of DBSCAN algorithm in the GIS uses the program code that was developed by Maneck and it is available on Github (Maneck 2014). Spatial data that used to be processed in this GIS using the DBSCAN algorithm is latitude and longitude. Clustering hotspot data was performed based on year, month, and province. Clustering parameters are epsilon and minimum points (MinPts). Epsilon values used range from 0.01 to 0.1 while MinPts ranges from 1 to 6. The clustering results are shown in form of table which consists of attribute Province, Regency, Latitude, Longitude, and Cluster.

Integrate module clustering and visualize hotspot

The GIS for hotspot distribution was developed using the client-server architecture. Clustering module as a server and visualizing hotspot in map as a client, because map can be shown by users. Software for clustering visualization module are MapServer and XAMPP v3.2.1 as a web-server. MapServer is a common gateway interface program that executed on a web server based on file in *.map. Map files save all definitions and configurations which needed by MapServer. File which is executed will bring out data to web browser. The map module in this GIS is integrated with GoogleMap (Figure 3 and Figure 4). Visualizing in map aims to show the result of clustering. The resulted clusters are shown with different colors. Different color represents each cluster.



Figure 3. Visualization of distribution hotspot based on clustering result

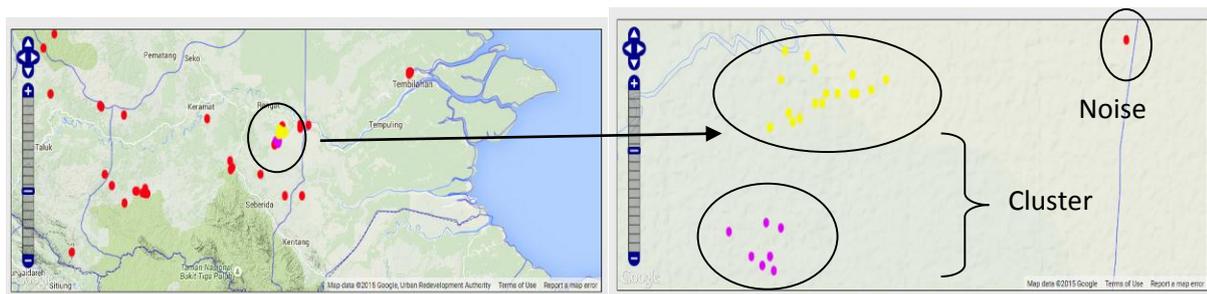


Figure 4. Visualization of hotspot clusters in zoom mode

CONCLUSION

The clustering module using the DBSCAN algorithm with PHP programming language is successfully built. Clustering results are shown in tables which consist of attribute Province, Regency, Latitude, Longitude, and Cluster. Map visualization for hotspots distribution based on clustering results is successfully developed using MapServer. The map module in this GIS is integrated with GoogleMap. The clustering module in GIS provides important information regarding distribution of hotspot clusters in Indonesia that may be used by related parties for fire prevention.

REFERENCES

Amri K. 2015. A Geographic Information System for Hotspot Occurrences Classification in Riau Province Indonesia [internet]. [cited 2015 August 7]; 24(2015):127-131. Available from: <http://www.sciencedirect.com/science/article/pii/S1878029615000857>
doi:10.1016/j.proenv.2015.03.17

- Ester M, Kriegel HP, Sander J, and Xu X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining.
- Kantardzic M. 2011. Data Mining: Concepts, Model, Methods, and Algorithms (2nd ed). Wiley-IIEEE-Press.
- Maneck B. [Internet]. 2014. DBSCAN-clustering. GitHub.com; [updated 2014 March 21; cited 2014 June 29]. Available from: <https://github.com/bhavikm/DBSCAN-clustering/blob/master/dbscan.php>.
- Mulya RD. 2014. Sistem Informasi Persebaran Titik Panas di Indonesia Menggunakan OpenLayers dan GeoExplorer. In: Seminar Nasional Pengarusutamaan Lingkungan Dalam Pengelolaan Sumberdaya Alam : Tantangan Dalam Pembangunan Nasional. Proceedings: 2014 Nov-6;Bogor.
- Purba C, Nanggara SG, Ratriyono M, Isnenti A, Rosalina L, Sari NA, Meridian AH. 2014. Potret Keadaan Hutan Periode 2009 – 2013. Bogor: Forest Watch Indonesia.