Incremental Spatio Temporal Clustering Application on Hotspot Data

I S Sitanggang†, N Radiatun†, A A Nur Risal†

†Department of Computer Science, Faculty of Mathematics and Natural Sciences, IPB University (Bogor Agrivicultural University), Bogor, Indonesia

*Corresponding author: imas.sitanggang@ipb.ac.id

Abstract. Hotspot is one of indications for forest and land fires. Analysis of hotspot data needs to be done as an early warning activity to prevent the occurrence of forest and land fires. The previous study analyzed hotspot clusters using the incremental spatio-temporal density-based clustering (ST-DBSCAN) algorithm. Hotspots in a cluster are considered as strong indicator for forest and land fires. However, clustering of hotspots is implemented on the command line interface. Through this interface, users are required to execute manually the commands to perform the clustering task on the dataset. The disadvantage of command line interface is that the commands have to be typed precisely and mistypes will cause errors.

Therefore, this study aims to build a web-based application for clustering hotspot data using the incremental ST-DBSCAN module. This study uses hotspot dataset in the period 2014 to 2017. The application was developed using R programming language and Shiny framework. The method of Adaptive Software Development (ASD) was adopted in this study. The main functions of the application are incremental clustering on hotspot data and visualization of clustering result. The testing using the black box approach show that all features of the application work properly. Based on the result of usability assessment, the user satisfaction level reach 78.45% meaning that the application is quite easy to be used.

1. Introduction

Indonesia has a large forest area. The total area of forest in Indonesia in 2015 reached 98 664 842.72 ha including protected forest area of 29 673 382.37 ha, limited production forest area of 26 798 382.01 ha, production forest area of 29 250 783.10 ha, and convertible production forest covering an area of 12 942 295.24 ha [1]. At present, the area does not have a significant increase due to forest degradation such as forest and land fires. In 2015 Indonesia experienced forest and land fires covering 261 060.44 ha [1].

Fires that often occur in Indonesia are also found in peatland forests. Fires in peatland forests are more difficult to deal with compared to fires in non-peatland. This is due to the spread of fire that not only occurs in vegetation on peat, but also occurs in peat soils that are difficult to determine [2]. The largest distribution of tropical peat in Indonesia in 2010 was found in three large islands: Sumatra, Kalimantan and Papua reaching an area of around 14.9 million hectares, not including peatlands on other islands [3]. The function of peatlands is to conserve water resources, reduce flooding, prevent sea water intrusion, support biodiversity, and control climate [4].

Forest and peatland fires can be identified through hotspots. According to the Regulation of the Minister of Environment and Forestry No P.8/Menhk/Setjen/Kum.1/3/2018 concerning Procedure for
field checking hotspots and/or forest and land fire information, hotspots that occurred in cluster and sequences are considered as strong indicator of forest and land fires therefore those hotspots become priority to be checked on the field by the patrol team.

The spatio-temporal clustering using the ST-DBSCAN algorithm has implemented to identify the hotspot clusters [5]. One of the issues in processing databases containing time series data including hotspot data is periodic data updates. Incremental data mining can be applied to overcome this issue by minimizing number of data scans and calculation efforts due to the occurrence of new data. In the implementation of ST-DBSCAN, incremental process is applied to dynamic databases to update the initial clusters. Incremental process of ST-DBSCAN was implemented on the hotspots data from April to May 2015 and September to October 2015 using the parameters Eps1 = 0.1, Eps2 = 3, 7, 30 and Minpts = 5 [5]. The results are old cluster with addition of new members, new clusters and outliers. The incremental process of ST-DBSCAN proposed in the previous work was conducted manually using console on R programming language [5]. This study aims to develop a web-based application for incremental spatio-temporal clustering on hotspot data using the Shiny framework so that updating hotspot clusters can be performed interactively.

2. Method

The data used in this study are hotspots data for the period of 2014-2017 that were obtained from NASA MODIS FIRMS (https://firms.modaps.eosdis.nasa.gov). The attributes included in clustering are latitude, longitude, and the time of the hotspot’s occurrence.

Shiny framework was used in developing web-based application that are integrated with the data mining algorithms. An application for generating and visualizing the hotspot sequences was built with the study area of Sumatra and Kalimantan [6]. Hotspots in a sequence are hotspot that are occurred consecutively 2-5 days in the same locations. Those hotspots are considered as strong indicators for forest and land fires. The implementation of spatial clustering on hotspot data of peatland in Sumatra was done using the DBSCAN algorithm and Shiny framework [7]. Moreover, Shiny framework was also used to create an application for detecting outliers on hotspot data using the K-Means Algorithm [8]. A web-based classification application for forest fire data was developed using the shiny framework and the decision tree C5.0 algorithm [9]. The dataset used in this study consists of ten explanatory layers (physical, weather, and socioeconomic characteristics) and one target layer (hotspot or non-hotspot) [9].

Application development was conducted through steps of Adaptive Software Development (ASD) with two iterations. Those steps include speculation, collaboration, and learning. At the speculation step, the incremental process in clustering hotspot data proposed in Sitanggang et al [5] was analyzed (Figure 1). The result of this stage is list functions required to be available on the application. All functions are implemented at the collaboration stage using the R programming language and Shiny framework. Shiny is a package in the R programming language that is used to provide interactive data summary through a web browser. Shiny has a variety of widgets to quickly build a user interface [10]. In the learning step, system testing was conducted. Two approaches were used in testing step namely black box testing dan usability testing. All functions in the system were tested using Black box approach. Furthermore, the system development continues to the second iteration based on the recommendations from Black box testing.

Usability testing was performed to evaluate the application, to validate the design, and to identify potency of problem occurrence in design. This study uses the rule of ISO 9241-11 in usability testing which has three measures namely effectiveness, efficiency, and satisfaction. Task and scenario in usability testing were designed based on the functions of the system. This study adopts one-on-one usability testing approach by observing one by one respondent while they are doing the task. The tool used in the usability testing is Post Study System Usability Questionnaire (PSSUQ). The success rate of completing tasks and comments from the respondent during the testing are then used to formulate the recommendation for system improvement. The effectiveness of the overall respondents in doing the task is calculated using Equation 1 [11].
effectiveness = $\frac{\sum_{j=1}^{R} \sum_{i=1}^{N} n_{ij}}{NR} \times 100\%$ \hspace{1cm} (1)

where

$n_{ij}$: the result of task-i by respondent-j
$N$: number of task that is done by respondent
$R$: number of respondents

The PSSUQ questionnaire was used to assess the application usability based on satisfaction with a likert scale of 1-7. The formula for calculating the satisfaction level is as follows [12].

$\text{Satisfaction level} = \frac{\sum_{j=1}^{M} \sum_{i=1}^{N} n_{ij}}{7NM} \times 100\%$ \hspace{1cm} (2)

where

$N$: number of questions on PSSUQ
$M$: number of respondents
$7$: Maximum value of likert scale
$n_{ij}$: answer-i by respondent-j

**Figure 1.** Incremental process in clustering hotspot data [5]
3. Result and Discussion
The development of incremental spatio temporal clustering application on hotspot data refers to the incremental process on ST-DBSCAN algorithm which is proposed in [5]. In order to get output of incremental clustering on hotspot data, the application has four features: ST-DBSCAN Clustering, Update Cluster, Help, and About. Under the menu ST-DBSCAN clustering and Update Cluster there are four sub menus namely Data, Cluster, Plot of Cluster, and Summary Cluster. Figure 2 illustrates the design of main pages of the application.

The application was implemented using the R programming language and Shiny framework. The hotspot data are stored on the database managed by DBMS PostgreSQL. The incremental process in clustering hotspot using ST-DBSCAN was implemented and stored in the file server.R whereas the system interface was implemented and stored in the file ui.R.

This study uses three global parameters of the ST-DBSCAN algorithm namely maximum radius of the neighbourhood (Eps1), the distance parameter for non-spatial attributes (Eps2) and minimum number of objects in an Eps-neighbourhood of the objects (MinPts). Figure 3 shows the interface of Cluster Update menu. In the first sidebar, the user selects the hotspot cluster to be updated. Hotspot clusters are generated using the ST-DBSCAN algorithm. The best parameter values of ST-DBSCAN algorithm on hotspot data were determined based on the silhouette index. The experimental results show that clusters generated at the Eps1 in the range 0.01 to 0.2 and Eps2 in the range of 7 to 30 have the highest silhouette index of 0.71. The range of Minpts selected is 3 to 7 days. Those selected parameter values are implemented in the application.

Figure 4 shows the cluster visualization represented using Google Maps. Different colours of hotspot on the map indicate different clusters of hotspots. On this visualization, number of hotspot clusters is
limited to 30. Figure 5 illustrates clusters of hotspot data as the results of incremental spatio-temporal clustering algorithm in Sumatra in October 2015 at the parameter \( \text{eps1} = 0.1, \text{eps2} = 30, \) and \( \text{MinPts} = 5. \)

![Visualization of Cluster 1 - Cluster 30](image)

**Figure 4.** Clusters visualization on Google Maps

![Clusters visualization on Google Maps](image)

**Figure 5.** Clusters of hotspot data as the results of incremental spatio-temporal clustering algorithm in Sumatra in October 2015 at the parameter \( \text{eps1} = 0.1, \text{eps2} = 30, \) and \( \text{MinPts} = 5. \)

This study applied the Blackbox testing approach to evaluate functions implemented in the application. Those functions are upload file of hotspot dataset, update cluster using the incremental spatio-temporal clustering algorithm, cluster visualization on Google Maps, display summary of cluster, and download the results of clustering in spreadsheet format. The testing results show that all those functions are working properly.
Usability assessment of the application was carried out using a one-on-one approach with the same procedure for each respondent. The assessment begins with the developer informing the purpose of the assessment, explaining the application and its functions, and explaining the questionnaire that must be filled out by the respondent. Respondents are free to comment on their opinions for the application and the developer records the time that the respondent worked on during the test. After all the tasks were completed, respondents were asked to provide a satisfaction assessment for the application using the PSSUQ questionnaire.

Task success rate is the number of tasks successfully completed by the respondent. The average rate of effectiveness of the overall respondents in doing the task is 98.33%, the value was calculated using Equation 1 [11]. The average rate of effectiveness of each respondent is provided in Figure 6.

![Figure 6. Average rate of effectiveness of each respondent](image)

The PSSUQ questionnaire was used to assess the application usability based on satisfaction with a likert scale of 1-7. The questionnaire assesses the application usability with indicators: overall, system usefulness, information quality, and interface quality as shown in Figure 7. The higher the PSSUQ score, the better the satisfaction indicator value. Based on the assessment by the respondents, the average of satisfaction level above the median value which is 4. The average of each satisfaction indicator is as follows: overall 5.49, system usefulness 5.65, information quality 5.19, and interface quality 5.69. The overall indicator in percentage was calculated using the formula (2) which reach 78.45%. We conclude that in general respondents are quite satisfied in using the application.

![Figure 7. Satisfaction level for each indicator](image)

4. Conclusion
The web-based application for incremental spatio-temporal density clustering on hotspot data has successfully developed to manage cluster of hotspots when the new hotspot data are included in the application. Shiny framework enables the web-based application to be integrated with the clustering
modules. The application has two main features: ST-DBSCAN Clustering to perform spatio-temporal density clustering on hotspot dataset, update cluster to conduct the incremental spatio-temporal density clustering on hotspot data. Clusters of hotspots are visualized on tabular format and maps. The application was tested using the Blackbox approach and the results show that all functions in the application are working properly. In addition, based on the results of system usability testing, the percentage of effectiveness is 98.33% and the satisfaction level is 78.45% which indicate that the respondents are quite satisfied in using the web-based application for incremental spatio-temporal clustering on hotspot data.

References

[1] Kementerian Lingkungan Hidup dan Kehutanan 2016 Statistik Kementrian Lingkungan Hidup dan Kehutanan Tahun 2015 (in Bahasa) (Jakarta: KLHK)

[2] Adinugroho W C, Suryadiputra I N N, Saharjo B H, Siboro L 2005 Panduan Pengendalian Kebakaran Hutan dan Lahan Gambut. Proyek Climate Change, Forests and Peatlands in Indonesia. (in Bahasa) Bogor: Wetlands International-Indonesia Programme and Wildlife Habitat Canada

[3] Ritung S, Wahyunto K, Nugroho, Sukarman, Hikmatullah, Suparto, Tafakresnanto C 2011 Peta Lahan Gambut Indonesia Skala 1:250.000 (Indonesian peatland map at the scale 1:250.000) (in Bahasa) Balai Besar Penelitian dan Pengembangan Sumberdaya Lahan Pertanian Bogor Indonesia

[4] Wahyunto, Ritung S, Suparto, Subagio H 2005 Sebaran Gambut dan Kandungan Karbon di Sumatera dan Kalimantan Proyek Climate Change, Forests and Peatlands in Indonesia. (in Bahasa) Wetlands International Indonesia Programme and Wildlife Habitat Canada Bogor: Wetlands International – Indonesia Programme

[5] Sitanggang I S, Risal A A N, Syaufina L 2018 Incremental Clustering on Hotspot Data as Forest and Land Fires Indicator in Sumatra. IOP Conference Series: Earth and Environmental Science 187 (1) pp. 1-14

[6] Abriantini G, Sitanggang I S, Trisminingsih R 2017 Hotspot sequential pattern visualization in peatland of Sumatera and Kalimantan using shiny framework. IOP Conf. Series: Earth and Environmental Science 54. pp: 1-7

[7] Hermawati R, Sitanggang I S 2016 Web-Based clustering application using Shiny framework and DBSCAN algorithm for hotspots data in peatland in Sumatra Procedia Environmental Sciences 33 pp. 317 – 323

[8] Suci A M Y A, Sitanggang I S 2016 Web-Based Application for Outliers Detection on Hotspot Data Using K-Means Algorithm and Shiny Framework IOP Conference Series: Earth and Environmental Science 31 pp. 1-8

[9] Siknun G P, Sitanggang I S 2016 Web-based classification application for forest fire data using the shiny framework and the C5.0 algorithm Procedia Environmental Sciences 33 pp. 332 – 339

[10] Beeley C 2013 Web Application Development with R using Shiny Birmingham (UK): Packt Publishing

[11] Bevan N 1995 Measuring usability as quality of use. Software Quality Journal 4(2) pp. 115-130

[12] Lewis J R 1995 IBM Computer usability satisfaction questionnaires: psychometric evaluation and instructions for use International Journal of Human-Computer Interaction 7(1) pp. 57-78.