

Available online at <http://www.mecs-press.net/ijem>

## Hotspot Sequence Patterns with an Improvement in Spatial Feature

Imas Sukaesih Sitanggang<sup>a</sup>, Dewi Asiah Shofiana<sup>a\*</sup>, Boy Sandi Kristian Sihombing<sup>a</sup>

<sup>a</sup>*Department of Computer Science, Bogor Agricultural University, Jl. Meranti Wing 20 Level V Dramaga, Bogor, 16680, Indonesia*

Received: 25 March 2018; Accepted: 08 June 2018; Published: 08 November 2018

---

### Abstract

Forest fires in Sumatra and Kalimantan resulted in degradation of peatlands significantly. The strong indicator of forest and land fires including in peatland can be identified using hotspots which occurred consecutively in 2 to 5 days. The previous studies have been conducted in mining sequence patterns on hotspot datasets in Sumatra and Kalimantan. However, those studies applied the sequential pattern algorithms on the datasets containing temporal and rough spatial features. This study aims to generate sequence pattern of hotspot datasets using the SPADE algorithm with the improvement of the spatial feature. The study results in 892 1-frequent sequences and 28 2-frequent sequence patterns at the minimum support of 0.02%. A total of 484 hotspots were found from the 28 2-frequent sequence patterns, most of which were occurred in September to November 2014 and 2015. Central Kalimantan, Riau, and South Sumatra are the area where hotspots mostly occurred in 2014 and 2015. The visualization module for hotspot sequences was successfully developed in two iterations using the JavaScript.

**Index Terms:** Hotspot, peatland fire, sequential pattern mining, SPADE.

© 2018 Published by MECS Publisher. Selection and/or peer review under responsibility of the Research Association of Modern Education and Computer Science.

---

### 1. Introduction

Peatland has a crucial role in our ecosystem. Indonesia has 14.9 million hectares of peatland spread in Sumatra, Kalimantan, and Papua Island. Unfortunately, the occurrence of forest fires every year threatens the preservation of the forest. From 2000-2009, there was degradation of peatland area approximately 2 million hectares. The degradation triggers effort from various parties to control and prevent the fires.

The appearance of a hotspot indicates the occurrence of the forest and land fire in a region. Hotspot is an area that has a higher temperature than its surroundings. Hotspot data have time and location information that are

\* Corresponding author. Tel.: +62 251 8625584; fax: +62 251 8625584  
E-mail address: [d.a.shofiana@gmail.com](mailto:d.a.shofiana@gmail.com)

collected each day resulting in large data. These data need to be processed to produce useful information and can be referenced for decision making.

Data mining and statistical method can be used to analyze the hotspot data as one of forest and land fire indicators. The study by Shi, 2012 created a model for estimating occurrences of large forest fires using the statistical method. Data mining has many techniques in managing, processing and analyzing data, among which is sequential pattern mining. Sequential pattern mining is a technique to mine sequence patterns from a sequence dataset. This method can be applied on hotspot dataset in order to obtain information of hotspot sequential occurrence patterns, which strongly indicates the incidents of forest and land fire in an area.

Research related to sequential pattern mining was done by Nurulhaq and Sitanggang, 2015. This study generated sequence patterns from hotspot dataset of Riau Province within the period of 2000-2014 using the PrefixSpan algorithm. The study results essential sequence patterns with support value ranging from 1%-11%. This study concluded that three days are the hotspot occurrence interval for being a strong indicator of fire. In addition, Agustina and Sitanggang, 2015 did a similar research in the same year but using a different algorithm. They used the Clospan algorithm and also added weather factor. The study shows that hotspots frequently occurred in an area with the precipitation rate of 0.3/inch/hour and temperature of 29.44°C.

The previous study chopped the three decimal digits of hotspot locations into two decimal digits. The consequence is the location of the generated hotspot patterns is not precise. The value 0.01 of hotspot coordinate indicates real radius value of about 1.1132 km. Chopping the digit can decrease the precise of hotspot locations to be quite far from its actual. Based on that, this study aims to generate sequence patterns on hotspot datasets using the sequential pattern discovery using equivalence classes (SPADE) algorithm with the improvement of the spatial feature. The improvement itself is by keeping all decimal digits of hotspot coordinate without rounding or chopping it. The visualization module for hotspot sequences as the result of the SPADE algorithm was developed in this study.

There are many other algorithms to mine sequential patterns such as Generalized Sequential Pattern (GSP) and Prefix-Projected Sequential Pattern (PrefixSpan). Based on the study by Verma and Mehta, 2014, PrefixSpan is superior in computation and memory efficiency rather than SPADE and GSP, but the number of frequent sequences that can be found is not as many as SPADE and GSP. Therefore, this study applied the SPADE algorithm to mine more frequent sequences than PrefixSpan, but the computation rate is faster than GSP.

## **2. Research Method**

This study uses the hotspot dataset in Sumatra and Kalimantan from January 2014 to December 2015. The data were collected from the Fire Information for Resource Management System (FIRMS) National Aeronautics and Space Administration (NASA). In addition, the 2002 peatland maps of Sumatra and Kalimantan from Wetland International are also used for hotspot selection. Table 1 shows the attributes of hotspot dataset. This study was conducted in several steps including data preprocessing, sequential pattern mining, and developing visualization module for hotspot sequences.

### *2.1. Data Preprocessing*

Data preprocessing was conducted to prepare a task-relevant data as an input for sequential patterns mining. The dataset was divided into four parts: Sumatra 2014, Sumatra 2015, Kalimantan 2014, and Kalimantan 2015. Tasks held in data preprocessing are data selection and data transformation. Chopping the decimal digits of hotspot coordinate is not applied, which differentiate this study with the previous study. Selection is a process to get data in the study area for further analysis. The selected hotspot's attributes are longitude, latitude, acquisition date (acq\_date), acquisition time (acq\_time), brightness temperature, and confidence. Hotspot data used are those within the peatland area which are selected using the peatland map from Wetland International. The number of hotspot in each dataset after selection is given in Table 2.

Table 1. Attributes of Hotspot (source <http://earthdata.nasa.gov>)

No.	Attribute	Description
1	Longitude	Location coordinate of hotspot (°)
2	Latitude	Location coordinate of hotspot (°)
3	Brightness Temperature	Hotspot temperature in channel 21 and 22 (K)
4	Scan	Width of satellite image pixel
5	Track	Length of satellite image pixel
6	Acq_date	Acquired date of hotspot
7	Acq_time	Acquired time of hotspot
8	Satellite	Satellite used (Aqua, Tera)
9	Confidence	Hotspot quality (0–100%)
10	Version	5.0 = MODIS NASA-LANCE 5.1 = MODIS MODAPS-FIRMS
11	Bright_t31	Hotspot temperature in channel 31 (K)
12	Frp	Fire radiative vigor (MegaWatt)

Table 2. Attributes of Hotspot (source <http://earthdata.nasa.gov>)

Dataset	Number of hotspots before selection	Number of hotspots after selection
Sumatra 2014	51760	40631
Sumatra 2015	62709	34497
Kalimantan 2014	40876	32575
Kalimantan 2015	75565	32491

After data selection, the dataset was transformed into sequential data format as an input for the SPADE algorithm as shown in Fig 1. The columns from left to right are sequence ID (SID), date\_code or event ID (EID), size, and items. The items column will appear as many as the value in size column. EID represents the code of hotspot occurrence date. The EID code is represented in integer as the results of the UNIXTIME function. The items column represents one or more EID which indicates the sequences of hotspots occurrence. As an example, in Fig 1 there is SID '2' with EID '1444853600' meaning that there is only one hotspot occurrence, so the EID is written once in the items column. If there are two hotspot occurrences such as SID '2' with EID '1444780800', the EID will be written twice in the items column.

```

1 1425686400 1 1425686400
2 1444780800 2 1444780800 1444780800
2 1444953600 1 1444953600
3 1425686400 1 1425686400

```

Fig.1. Sequential Data Input for SPADE

## 2.2. Sequential Pattern Mining

Sequential pattern mining is a process to extract sequential patterns that its support exceeds the minimum support. The value of minimum support usually defines by users. This value will be a threshold for selecting sequence patterns, which patterns that are less attractive can be ignored so that the mining process will be more

efficient (Zaki, 2001).

The sequential pattern is the pattern acquired from itemset with its appearance is sequential, and all items happen at almost the same time (Zhao and Bhowmick, 2003). Sequence  $\langle a_1, a_2, \dots, a_n \rangle$  is also part of sequence  $\langle b_1, b_2, \dots, b_m \rangle$ , if  $i_1 < i_2 < \dots < i_n$  and  $a_1 \subseteq b_{i_1}, a_2 \subseteq b_{i_2}, \dots, a_n \subseteq b_{i_n}$ . For example, sequence  $\langle (C)(F, G, I)(G, I) \rangle$  is part of sequence  $\langle (B)(C)(F, G, H, I)(G, I) \rangle$  because  $(C) \subseteq (C), (F, G, I) \subseteq (F, G, H, I), (G, I) \subseteq (G, I)$ . But,  $\langle (B)(C) \rangle$  is not part of  $\langle (B, C) \rangle$  because  $\langle (B)(C) \rangle$  means C happens after B whereas  $\langle (B, C) \rangle$  means B happens at the same time with C. A sequence is called maximum if it is not a part of any other sequences (Zhao and Bhowmick, 2003).

Sequential pattern discovery using equivalent classes (SPADE) is an algorithm for mining frequent sequence patterns using the equivalence class by splitting problems into sub problems which can be solved separately and then conquer it using the combining operation (Zaki, 2001). General steps of the SPADE algorithm (Zaki, 2001):

1. Count the support value of all items
2. Count the support value of all  $k$ -frequent sequence items, where  $k$  is the maximum length of the sequence that its support exceeds the minimum support
3. Decompose a class of all sequences based on its length
4. Enumerate the entire sequences to generate new sequence

A sequence pattern is a list describing the appearance sequence of items in a dataset. Hotspots sequence patterns were obtained from the transformed data in the previous step. Sequence data were processed using the SPADE Algorithm which is available in R within the 'arulesSequence' package.

### 2.3. Developing Visualization Module

The application for hotspot sequence visualization was developed by adopting the steps in *Adaptive Software Development* (ASD) (Pressman, 2010). Software used to develop the application are Java programming language for creating web pages, the library SPMF (Fournier-Viger et al., 2014) to implement the SPADE algorithm, DBMS MySQL and JavaScript for visualization. The steps in ASD include speculation phase, collaboration phase, and learning phase.

Several activities performed in the speculation phase are understanding the data used, purposes of application development, and data preparation (Alnoukari et al., 2008). In the collaboration phase, modeling subsections of the application were designed through discussions conducted with potential users. Modeling was done by designing the features in the application, identifying the priority level of each feature, and determining the duration of the application development. In the learning phase, the feature that has been created in each iteration was tested. Testing scenario includes data test, application test, and exceptions. Responses from the testing were presented through the sharing session to solve the constraints encountered.

## 3. Result and Discussion

The research results sequence patterns generated by the SPADE algorithm for each dataset. Only some interesting generated patterns will be selected for further analysis.

### 3.1. Generate and Select Sequence Patterns

Sequence patterns were generated in R using the `cspade` function which is available in `arulesSequence` package. Minimum support is set to 0.02% to generate 2-frequent sequences for all datasets. A sequence with two items is called 2-frequent sequence. Fig. 2 shows the example of SPADE's output. The number of sequences generated is given in Table 3.

	sequence	support
134	<{1448582400}>	0.0002213089
135	<{1448755200}>	0.0002529244
136	<{1448928000}>	0.0009800822
137	<{1449532800}>	0.0003161555
138	<{1450915200}>	0.0004742333
139	<{1451260800}>	0.0002529244
140	<{1444780800}, {1444953600}>	0.0006006955
141	<{1444348800}, {1444780800}>	0.0002529244
142	<{1442793600}, {1442966400}>	0.0003161555
143	<{1442188800}, {1442620800}>	0.0002213089
144	<{1442016000}, {1442188800}>	0.0002845400

Fig.2. Example of SPADE's Output

Table 3. Number of Sequences Generated by SPADE

Dataset	1-frequent sequence	2-frequent sequence
Sumatra 2014	241	6
Sumatra 2015	226	12
Kalimantan 2014	286	4
Kalimantan 2015	139	6

The selected sequence patterns were only patterns with minimum two hotspot occurrences and the length of 2-sequence. Each selected sequence patterns were stored in CSV file to identify the location and other attributes of the hotspot. As shown on Fig. 2, there is a sequence <{1444780800}, {1444953600}> with support value of 0.0006006955. Date\_code 1444780800 and 1444953600 sequentially refers to October 14th, 2015 and October 16th, 2015. Based on the sequence, it can be interpreted that on October 14th, 2015 there were hotspots in some locations that occurred again on October 16th, 2015.

### 3.2. Sequence patterns of Sumatra 2014 and 2015

Table 4 and Table 5 show the sequences generated from the datasets of Sumatra. The plot of sequence patterns using QuantumGIS is provided in Fig. 3 and Fig. 4. Based on Table 4 and Table 5, it is shown that hotspots mostly occurred around February to March in 2014 and September to October in 2015. The support value infers to the appearance of sequence whereas formed as a date\_code in the dataset. Fig. 4 shows that Riau and South Sumatra are two provinces where hotspot occurrence mostly detected. Most hotspots in Riau occurred in Bengkalis, whereas Ogan Komering Ilir has the highest rate of hotspot occurrence in South Sumatra.

Table 4. 2-frequent Hotspot Sequence Patterns of Sumatra 2014

Sequence ID	Date_code	Date	Support
242	<1395705600, 1395878400>	<25 Mar '14, 27 Mar '14>	0.000227043
243	<1394496000, 1394668800>	<11 Mar '14, 13 Mar '14>	0.000302725
244	<1394323200, 1394496000>	<9 Mar '14, 11 Mar '14>	0.000302725
245	<1393113600, 1393545600>	<23 Feb '14, 28 Feb '14>	0.000201816
246	<1393459200, 1393545600>	<27 Feb '14, 28 Feb '14>	0.000201816
247	<1392940800, 1393113600>	<21 Feb '14, 23 Feb '14>	0.000201816

Table 5. 2-frequent Hotspot Sequence Patterns of Sumatra 2015

Sequence ID	Date_code	Date	Support
227	<1445644800, 1445817600>	<24 Okt '15, 26 Okt '15>	0.00027243
228	<1445385600, 1445644800>	<21 Okt '15, 24 Okt '15>	0.00039351
229	<1445558400, 1445644800>	<23 Okt '15, 24 Okt '15>	0.00021189
230	<1445212800, 1445472000>	<19 Okt '15, 22 Okt '15>	0.00024216
231	<1445385600, 1445472000>	<21 Okt '15, 22 Okt '15>	0.00030270
232	<1445212800, 1445385600>	<19 Okt '15, 21 Okt '15>	0.00024216
233	<1444003200, 1444262400>	<5 Okt '15, 8 Okt '15>	0.00021189
234	<1443657600, 1443830400>	<1 Okt '15, 3 Okt '15>	0.00021189
235	<1442880000, 1443225600>	<22 Sep '15, 26 Sep '15>	0.00021189
236	<1443052800, 1443225600>	<24 Sep '15, 26 Sep '15>	0.00024216
237	<1442880000, 1443052800>	<22 Sep '15, 24 Sep '15>	0.00036324
238	<1441238400, 1441324800>	<3 Sep '15, 4 Sep '15>	0.00024216

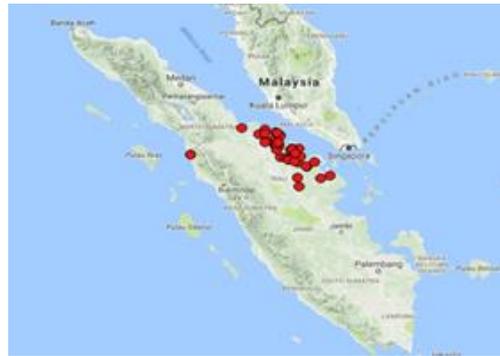


Fig.3. Plot of Sequences in Sumatra 2014

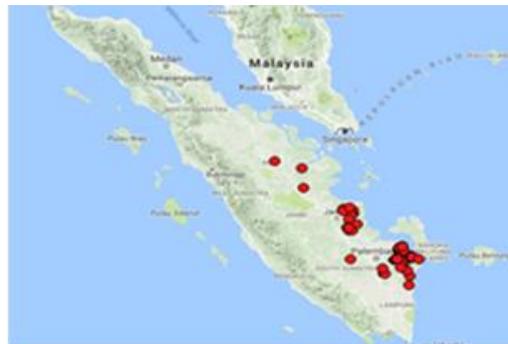


Fig.4. Plot of Sequences in Sumatra 2015

### 3.3. Sequence Patterns of Kalimantan 2014 and 2015

Table 6 and Table 7 show the sequences generated from the datasets of Kalimantan. The sequence patterns

plotted using Quantum GIS shown in Fig. 5 and Fig. 6. Based on Table 6 and Table 7, it is shown that hotspots mostly occurred from September to November. Fig. 6 and Fig. 7 show Central Kalimantan as a high potential area of hotspot emergence, especially in Katingan and Pulang Pisau district.

Table 6. 2-frequent hotspot Sequence Patterns of Kalimantan 2014

Sequence ID	Date_code	Date	Support
287	<1414800000, 1414972800>	<1 Nov '14, 3 Nov '14>	0.000252924
288	<1412726400, 1412812800>	<8 Okt '14, 9 Okt '14>	0.000242160
289	<1412035200, 1412208000>	<30 Sep '14, 2 Okt '14>	0.000211890
290	<1411603200, 1411776000>	<25 Sep '14, 27 Sep '14>	0.000211890

Table 7. 2-frequent Hotspot Sequence Patterns of Kalimantan 2015

Sequence ID	Date_code	Date	Support
140	<1444780800, 1444953600>	<14 Okt '15, 16 Okt '15>	0.000600696
141	<1444348800, 1444780800>	<9 Okt '15, 14 Okt '15>	0.000252924
142	<1442793600, 1442966400>	<21 Sep '15, 23 Sep '15>	0.000316155
143	<1442188800, 1442620800>	<14 Sep '15, 19 Sep '15>	0.000221309
144	<1442016000, 1442188800>	<12 Sep '15, 14 Sep '15>	0.000284540
145	<1441670400, 1441756800>	<8 Sep '15, 9 Sep '15>	0.000221309



Fig.5. Plot of Sequences in Kalimantan 2014



Fig.6. Plot of Sequences in Kalimantan 2015

### 3.4. Comparison with Previous Study

The identified hotspots from the sequence pattern were compared with hotspots from the previous study of Abriantini et al., 2017. The purpose of comparison was to observe the location's conformity of 2-frequent sequences from this study and the previous one. The comparison was made by rounding the location coordinates into two decimal digits and find the hotspot's locations that also exist in the previous study by Abriantini et al., 2017. The comparison shows that some locations of hotspot were not identified in the sequence pattern from the previous study. In the study by Abriantini et al., 2017, the rounding of location coordinates was done at the very first step, and only hotspots with confidence above 70% were used. Besides, the minimum support determined in the previous study was 0.01.

The comparison visually is shown in Fig. 7, Fig. 8, and Fig. 9. Fig. 7 shows the hotspots sequences from this study overlaid with the sequences from Abriantini et al., 2017 in a wide-scale scene and is hardly seen. Fig 8 shows the sequences from both studies in a zoomed-in scene with the red points refers to the sequences from this study and the green points from Abriantini et al., 2017. Fig. 9 shows the sequences in a closer view so it can be seen that there is quite a distance difference between the sequences of both studies.

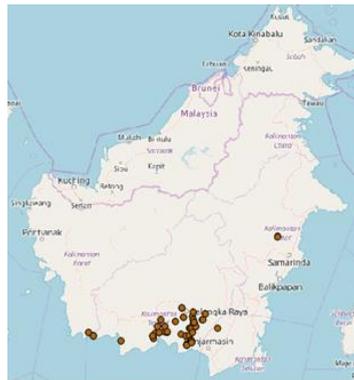


Fig.7. Sequences in Kalimantan 2015 Overlaid with the Result of the Study by Abriantini et al., 2017



Fig.8. Zoomed-in view of Sequences in Kalimantan 2015 Overlaid with the Result of the Study by Abriantini et al., 2017

When the hotspots from sequence patterns in the Abriantini et al., 2017 were verified to the field, the locations were not precise resulting longer time of verification. The reason could be the rounding into two

decimal digits at the beginning, affecting location error for about 1 km. Fig. 10 and Fig. 11 shows the distance measurement of selected hotspots from both studies.



Fig.9. Closer Zoomed-in view of Sequences in Kalimantan 2015 Overlaid with the Result of the Study by Abriantini et al., 2017

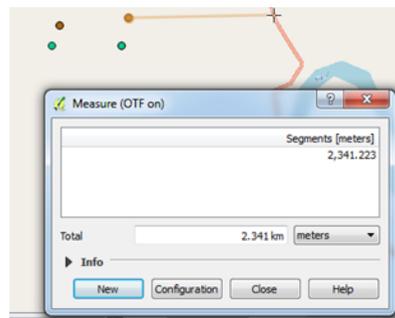


Fig.10. Distance Measurement of Selected Hotspot from This Study

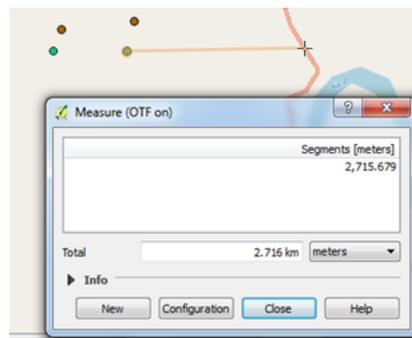


Fig.11. Distance Measurement of Selected Hotspot from the Study by Abriantini et al., 2017

Fig. 10 and Fig. 11 show the distance measurement of the selected hotspot to the main road. Closer to the main road is better because it is easier and faster for the officials in detecting and extinguishing the fire, meaning less area affected. From Fig. 10 and Fig. 11 it is shown that the distance from this study is 2.341 km whereas the previous study by Abriantini et al., 2017 is 2.761 km, making 420 meters difference or nearly 0.5 km. With the topology of peatland forest, a distance of 0.5 km makes quite a big difference in time for the officials in extinguishing the fire. From the comparison, it is shown that by keeping the three decimal digits of coordinates, hotspots location from this study is more precise rather than most previous studies.

### 3.5. Visualization Application for Hotspot Sequences

The application for visualization hotspot sequences was developed in two iterations. Each iteration adopts the steps in ASD namely speculation phase, collaboration phase, and learning phase. The main objective of the first iteration is to create a module for generating hotspot sequence pattern using the SPADE algorithm which is available in the SPMF tool.

In the speculation phase, the system requirements are collected. User's needs were obtained from 13 respondents who are Fire Brigade in the operational region of Banyuasin, Palembang, South Sumatra. The user's needs were collected on 21 August 2017. Features of the application required by the users are as follows:

1. The application is expected to be accurate and factual in visualizing the location of the hotspot sequence patterns. This means that the locations of the hotspot sequence patterns are visualized in the form of a map.
2. The application is expected to display the process of sequences generation in the format that is easy to be read and to be understood by the users. This means that each result of the data processing will be displayed in the tabular format.
3. The application is expected to be easily used. This means there is a feature that provides the guideline about the procedure of using the application.

In the collaboration phase a database was designed. There are four tables in the database which store hotspot datasets, hotspot data as the result of pre-processing, and sequences of hotspot. Mockup of the application was created using the Java programming language and JavaScript. The results of the collaboration phase are web pages and its features including data pre-processing, data transformation, sequences generation, and sequence pattern visualization. Fig. 12 shows the visualization of hotspot sequences as the result of the first iteration.



Fig.12. Visualization of Hotspot Sequences as the Result of the First Iteration

In the learning phase, the system was tested to evaluate the output of the features. The testing results show that the features data pre-processing, data transformation, sequences generation, and sequence pattern visualization are working properly.

The second iteration focuses on creating the visualization module for hotspot sequences as output of the SPADE algorithm which is implemented in the SPMF tool. In the speculation phase, we identified the

functions of the visualization module. The hotspot sequences which are generated using the SPADE algorithm are classified into several groups based on the length of sequences. In this study the algorithm results 1-frequent sequences, 2-frequent sequences, 3-frequent sequences, and 4-frequent sequences. The characteristics of sequence are then selected from the database. Those characteristics include locations of the sequence represented in latitude, longitude, date, province, district, sub district and village where the hotspot sequences are located. The design and mockup of the visualization module were realized in the collaboration phase. The implementation of module was done using JavaScript Application Program Interface (API). The system testing was performed in the learning phase. The testing result shows that all features in the visualization module are working properly. Fig. 13 shows the visualization of 1-frequent hotspot sequences as the result of the second iteration. The visualization of 2-frequent hotspot sequences and 4-frequent hotspot sequences are illustrated in Fig. 14 and Fig. 15 respectively.



Fig.13. Visualization of 1-frequent Hotspot Sequences



Fig.14. Visualization of 2-frequent Hotspot Sequences



Fig.15. Visualization of 4-frequent Hotspot Sequences

#### 4. Summary

There are 6 2-frequent sequence patterns of Sumatra 2014, 12 2-frequent sequence patterns of Sumatra 2015, 4 2-frequent sequence patterns of Kalimantan 2014, and 6 2-frequent sequence patterns of Kalimantan 2015 resulted from SPADE Algorithm at the minimum support of 0.02%. 484 hotspots were found from 28 2-frequent sequence patterns which mostly occurred in Central Kalimantan, Riau, and South Sumatra from September to November 2014 and 2015. The hotspots locations within the generated sequences are more precise than previous studies so it will be helpful for the officials in detecting and extinguishing the fire. The visualization module of hotspot sequences will help users to obtain the location of hotspot sequences which is important for forest and land fire prevention.

#### References

- [1] Abriantini G, Sitanggang IS, Trisminingsih R. Hotspot sequential pattern visualization in peatland of Sumatera and Kalimantan using Shiny framework. *IOP Conference Series: Earth and Environmental Science* 2017; 54:1-7.
- [2] Agustina T, Sitanggang IS. Sequential patterns for hotspot occurrences based weather data using Clospan algorithm. *International Conference on Adaptive and Intelligent Agroindustry (ICAIA) 2015*; 3:301-305.
- [3] Alnoukari M, Alzoabi Z, Hanna S. Applying adaptive software development (ASD) agile modelling on predictive data mining applications: ASD-DM methodology. *IEEE Proceedings of International Symposium on Information Technology* 2008; 1083-1087.
- [4] Fournier-Viger P, Gomariz A, Gueniche T, Soltani A, Wu CW, Tseng VS. SPMF: A Java Open-Source Pattern Mining Library 2014; available at <http://www.jmlr.org/papers/volume15/fournierviger14a/fournierviger14a.pdf>
- [5] Nurulhaq NZ, Sitanggang IS. Sequential pattern mining on hotspot data in Riau Province using the PrefixSpan algorithm. *International Conference on Adaptive and Intelligent Agro-industry (ICAIA) 2015*;3:257-260.
- [6] Pressman RS. *Software engineering: A practitioner's approach*. 7th ed. New York: McGraw-Hill; 2010.
- [7] Shi Y. A probability model for occurrences of large forest fires. *IJ. Engineering and Manufacturing* 2012; 2(1):1-7.
- [8] Verma M, Mehta D. Sequential pattern mining: A comparison between GSP, SPADE, and PrefixSpan. *International Journal of Engineering Development and Research (IJEDR)* 2014; 3016-3036.
- [9] Zaki MJ. SPADE: An efficient algorithm for mining frequent sequences. *Machine Learning* 2001; 42:31-60.
- [10] Zhao Q, Bhowmick SS. *Sequential pattern mining: A survey*. ITechnical Report CAIS Nayang Technological University Singapore 2003; 1-26.

#### Authors' Profiles



**Imas Sukaesih Sitanggang** is a lecturer in Computer Science Department, Bogor Agricultural University, Indonesia. Her main research interests include spatial data mining and data warehousing.



**Dewi Asiah Shofiana** is a master student of Computer Science in Bogor Agricultural University, Indonesia. Her main research interest is spatial data mining.

**Boy Sandi Kristian Sihombing** is a undergraduate student of Computer Science in Bogor Agricultural University, Indonesia. His main research interest is data mining.

**How to cite this paper:** Imas Sukaesih Sitanggang, Dewi Asiah Shofiana, Boy Sandi Kristian Sihombing, "Hotspot Sequence Patterns with an Improvement in Spatial Feature", International Journal of Engineering and Manufacturing(IJEM), Vol.8, No.6, pp.13-25, 2018.DOI: 10.5815/ijem.2018.06.02