Available online at www.sciencedirect.com

ScienceDirect

Procedia Environmental Sciences 33 (2016) 324 – 331

Procedia Environmental Sciences

The 2nd International Symposium on LAPAN-IPB Satellite for Food Security and Environmental Monitoring 2015, LISAT-FSEM 2015

Challenges in satellite-based research on forest and land fires in Indonesia: frequent item set approach

Lailan Syaufina[a,]*, Imas Sukaesih Sitanggang[b], Lusi Maulana Erman[b]

[a]*Faculty of Forestry, Bogor Agricultural University, Jl. Lingkar Akademik Kampus IPB Darmaga, Bogor 16680, Indonesia*
[b] *Faculty of Mathematics and Natural Sciences,, Bogor Agricultural University, Jl. Meranti Kampus IPB Darmaga, Bogor 16680, Indonesia*

Abstract

Forest fire research in Indonesia has been in increasing trend since 1982/1983- and 1997/1998-fire episodes. The first episode emphasized on fire impacts and the second episode has more research on emission and pollution. Satellite-based researches on fire are scattered and still limited. This study aimed to analyze fire satellite-based research aspects and challenges for the future. Text mining was applied to analyze selected international as well as national journals containing fire satellite-based research. The study revealed that there are three main research clusters: fire emission and pollution, fire detection, and fire danger and risk.

Peer-review under responsibility of the organizing committee of LISAT-FSEM2015

Keywords: forest fire; fire episode; satellite-based research; emission; text mining; Indonesia.

1. Introduction

One of the important environmental problems in Indonesia is forest and land fire occurrences, which has been pronounced since 1982/1983 when about 3.6 million ha of tropical rain forest in East Kalimantan burned out. Large fire events recurrent in 1987, 1990/1991, 1994, 1997/1998, 2002, 2006, 2013, 2014, and currently in 2015, which correspond to extreme weather events such as El Nino phenomenon. Two years of satellite-based active fire detections over Peninsular Malaysia, Sumatra, Borneo and Java were examined together with land cover and peat

* Corresponding author. Tel.: +62-813-1102-8233.
E-mail address: syaufina2016@gmail.com.

Peer-review under responsibility of the organizing committee of LISAT-FSEM2015
doi:10.1016/j.proenv.2016.03.083

land maps [1] showed that fire occurrence nearly tripled (23,000 vs 68,000) from a wet La Niña year (2008) to a drier El Niño year (2009). In both years, fires were concentrated in peat lands (in 2009 41% of fires vs.10% of land area), and the majority of large-scale burning took place in peat lands. Impacts of the fires were experienced by not only local level but also global environment. For science, those fire episodes have been a great challenge to study the impacts and to develop methodology of assessment as well as fire control.

Remote sensing as satellite-based technology plays very important role in forest and land fire assessment, which include: fire detection, fire impact assessment, early warning system, fire management plan, and post fire rehabilitation plan. Furthermore, satellite data have been used to monitor biomass burning at regional and global scale for more than two decades using algorithms that detect the location of active fires at the time of satellite overpass and in the last decade using burned area algorithms that map directly the spatial extent of the area affected by fires [2] such as the MODIS Burned Area Product (MCD45). Forest fire research is one of many appropriate GIS applications. The diversity of factors that affect the beginning and spreading of a forest fire dictates the use of an integrated analysis approach. Considering the intrinsic dynamism of this phenomenon, remote sensing imagery is also very valuable for these kinds of studies. It provides a quick evaluation of the vegetation status, as well as a survey of the effects of fire upon the environment [3]. However, information on the extent and magnitude of the technology application is still limited.

On the other hand, text mining4 commonly defined as a knowledge-intensive process in which a user interacts with a document collection over time by using a suite of analysis tools. It is a new and exciting research area that tries to solve the information overload problem by using techniques from data mining, machine learning, natural language processing (NLP), information retrieval (IR), and knowledge management. The data sources in text mining are document collections, interesting patterns are found not among formalized database records but in the unstructured textual data in the documents in these collections. The technology has been now widely applied including in business, research, and government needs such as biomedical application, software application, online media applications, marketing application, and academic application for examples.

Therefore, this paper aimed to elaborate how far the satellite-based research has been developed and to what extent the research were conducted through text mining analyses using frequent item set approach of journal articles on fire satellite-based researches. The analyses results will be used to identify the challenge of satellite-based research on forest and land fires to be developed in the future, which will provide benefit to minimize forest and land fire occurrences in Indonesia.

2. Data and Methods

2.1. Study area

This study is covering satellite-based researches on forest and land fires conducted in Indonesia as well as those related to fire occurrences and the impacts in Indonesia, particularly in Sumatera and Kalimantan (Fig. 1).

Fig. 1. Map of Indonesia

2.2 Data and tools

Datasets used in the study were 76 published as the free available research articles published on satellite-based research on forest and land fires in Indonesia that were collected from several sources in national as well as international journals. The articles were in PDF and Microsoft Word format.

3. Methods

The articles in PDF and Microsoft Word format were converted to Text files in order to be processed in the documents clustering using the K-means algorithm. Three main steps including 1) data preprocessing, 2) generating frequent itemsets, and 3) documents clustering were conducted to analyze the articles represented in Text files. We performed several steps in data preprocessing to prepare a task relevant dataset for generating frequent itemsets on text documents. These steps are case folding, removing punctuations and numbers, filtering, deleting whitespace, stemming, and creating a document-term matrix. In case folding, all letters in documents are presented on lowercase [4]. Stopwords that are common words with have no analytic value are removed in filtering step. Stemming is the process to remove all affixes in the words [4]. The last step in data preprocessing is creating a document-term matrix that contains the frequency of terms occurrence in the documents. Output of data preprocessing is a dataset as input of the algorithm for generating frequent item sets.

Frequent itemsets are collection of items that are frequently occurred in a transactional dataset that have support greater than the user-specified minimum support (*minsup*). Discovering frequent itemsets is the first task in association rule mining in addition to rule generating from the frequent itemsets. Several algorithms can be applied in association rule mining including Apriori [5], Frequent Pattern Growth (FP-Growth) [6] and Equivalence Class Transformation (ECLAT). Association rule mining discovers frequent itemsets and association rules in a transactional dataset that have support and confidence greater than the user-specified minimum support (*minsup*) and minimum confidence (*minconf*) respectively. An association rule has the form $A \rightarrow B$, where A and B are a subset I, I is a set of items, and $A \cap B = \emptyset$. *Support* ($A \rightarrow B$) is the percentage of transactions in a dataset D that contain both A and B. *confidence*($A \rightarrow B$) is the percentage of transactions in D containing A that also contain B [6]. Support and confidence of the rule $A \rightarrow B$ are defined as follows [6]:

$$support(A \rightarrow B) = P(A \cup B) \tag{1}$$

$$confidence(A \rightarrow B) = P(B \mid A) = \frac{support(A \cup B)}{support(A)} \tag{2}$$

In this work, we applied the ECLAT algorithm to generate frequent itemsets from a collection of documents in which terms in the documents are represented in a vertical layout. The ECLAT algorithm is as follows where the term atom in the algorithm represents a term in a document [7] (Table 1.):

The ECLAT algorithm discovered terms that are frequently appear in the documents. Frequency of term occurrence in each document was calculated and the result was stored in the document-term matrix in which columns in the matrix represent frequent terms and rows in the matrix represent document id. The document-term matrix is the input for the clustering algorithm to group documents based on the frequent terms. Clustering was performed using the widely used algorithm namely K-Means that is available in the programming language R. The main tasks in the K-Means algorithm are as follows [6]:

1 Select randomly k objects in the dataset as the initial centers of clusters. The centers are called centroids.
2 Repeat until clusters do not change
   a (re)assign each object into the cluster to which the object is the most similar, based on the mean value of objects in the cluster;
   b Update the *centroid* value of the *cluster*, by calculating the mean value of the objects for each cluster.

Sum of squared error (SSE) is used to determine the best value of number of cluster (k). Clustering with the smallest SSE is the best clustering result. SSE is defined as follows[6]:

$$\mathrm{SSE} = \sum_{i=1}^{k} \sum_{x \in C_i} \mathrm{dist}(c_i, p)^2 \quad (3)$$

where k is the number of clusters, p is a data object, C_i are objects in the cluster i, c_i is the centroid or center point of the cluster i, and *dist* is the distance function, for example Euclidean distance.

Table 1. The ECLAT algorithm

Algorithm 1: ECLAT – Frequent Itemset Mining	
Input:	A transaction database D, A user specified threshold min_{sup} A set of atoms of a sublattice S
Output:	Frequent itemsets F

Procedure:
Eclat(S)
	for all atoms $A_i \in S$
	$T_i=\emptyset$
	for all atoms $A_j \in S$, with $j > i$ *do*
	$R= A_i \cup A_j$;
	$L(R)=L(A_i) \cap L(A_j)$;
	If $support(R) \geq min_{sup}$ *then*
	$T_i=T_i \cup \{R\}$;
	$F_{\|R\|}=F_{\|R\|} \cup \{R\}$;
	end
	end
	end
	for all $T_i \neq \emptyset$ do Eclat(T_i);

4. Result and Discussion

4.1 Temporal distribution of articles of forest and land fires in Indonesia

Research on forest and land fires have been on the increasing trend since 1982/1983 [8] as well as those related to peatland fire [9]. Similarly, fire satellite-based research has indicated positif and increasing trend since year of 2000 and has significantly increased since 2012 (Fig.2). The number of articles ranged from 0 (2003) to 13 (2015) with 5 article number per year in average.

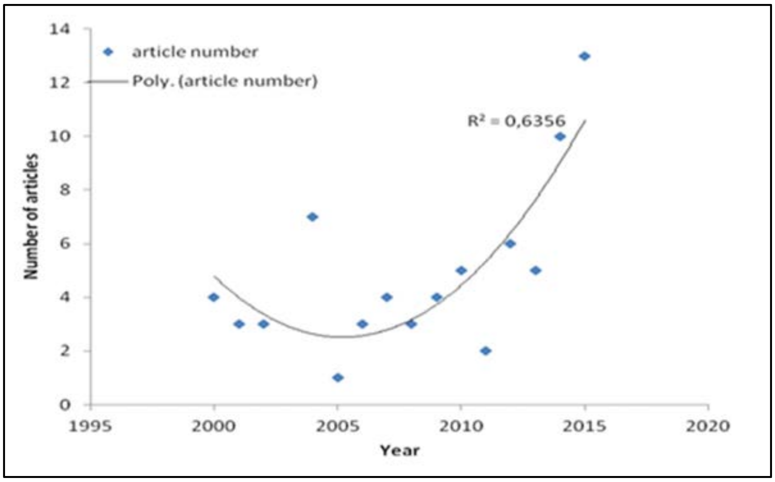

Fig. 2. Distribution of number of articles on forest and land fires satellite-based research.

The tremendous increase in the period of 2012-2015 seems to be influenced by fire occurrences and related impacts of haze in particular,which covered Sumatera and Kalimantan. For what extent the research address the fire problem, further analyses need to be performed.

4.2 Cluster distribution on articles of satellite-based research on forest and land fires in Indonesia

Applying text mining approach, the number of articles can be classified into 10 clusters with various SSE (Fig. 3). The highest SSE (15792.74) is found in Cluster 01, whereas the lowest SSE (10150.14) is found in Cluster 10, which means that number of articles is classified into 10 fire aspects, namely: fire emission, fire danger, fire detection, fire impacts assessment, fire pollution, fire risk mapping, fire hotspot indicator, fire-drought, fire burned area, and fire haze/smoke.

```
Cluster = 01 | SSE= 15792.74 | Ukuran tiap cluster=76
Cluster = 02 | SSE= 15045.55 | Ukuran tiap cluster=63 13
Cluster = 03 | SSE= 14089.13 | Ukuran tiap cluster=52 20 04
Cluster = 04 | SSE= 13164.12 | Ukuran tiap cluster=52 19 04 01
Cluster = 05 | SSE= 13164.12 | Ukuran tiap cluster=04 05 03 01 63
Cluster = 06 | SSE= 12671.69 | Ukuran tiap cluster=11 03 04 01 55 02
Cluster = 07 | SSE= 11995.78 | Ukuran tiap cluster=11 03 03 01 55 01 02
Cluster = 08 | SSE= 11413.85 | Ukuran tiap cluster=40 05 01 01 22 01 02 04
Cluster = 09 | SSE= 10732.27 | Ukuran tiap cluster=03 03 03 01 51 01 02 01 1
Cluster = 10 | SSE= 10150.14 | Ukuran tiap cluster=03 01 03 01 47 01 02 01 1
```

Fig. 3. Cluster of forest and land fire satellite-based research in Indonesia

Previous studies [8] indicated that the 1982-1983-fire episode promoted research on fire impacts on biodiversity, particularly in East Kalimantan, where the largest forest fire event occurred in the region. The 1997/1998-fire episode has promoted two important aspects of study, namely: fire impacts on biodiversity and fire impacts on human dimensions. In the period of 2000-2015, the fire satellite-based research aspects seem to be improved in scope and magnitude.

Further content analyses of the articles revealed that those 10 fire aspects can be merged into 6 aspects (cluster 06), namely: fire emission, fire detection, fire impacts assessment, fire pollutant, fire risk, and fire indicator (Fig. 4). Fire emission ranks the highest number of articles, followed by fire detection and fire risk, fire pollution, fire impacts, and fire indicator.

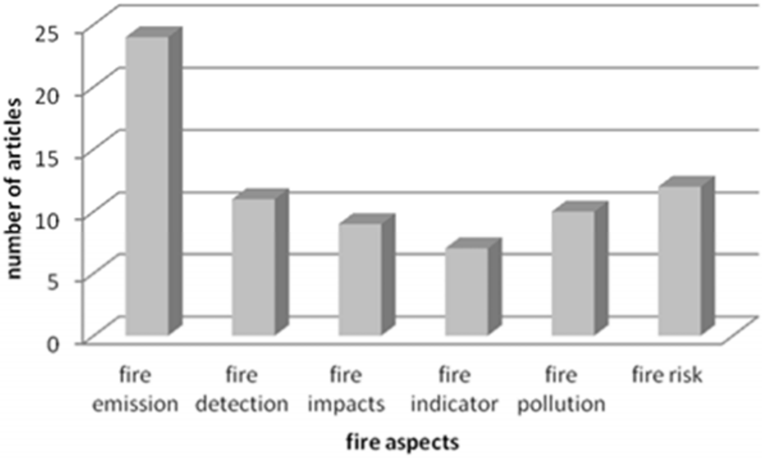

Fig. 4. Cluster 06 of forest and land fire satellite-based research in Indonesia

By deepening further content analyses revealed that the fire satellite-based research may only categorized in three aspects (cluster 03) as the most likely cluster, namely: fire emission and pollution, fire risk and detection, and fire impacts assessment (Fig. 5). The highest number of research articles found in fire emission and pollution (46%), followed by fire risk and detection (42%), and fire impacts assessment (12%).

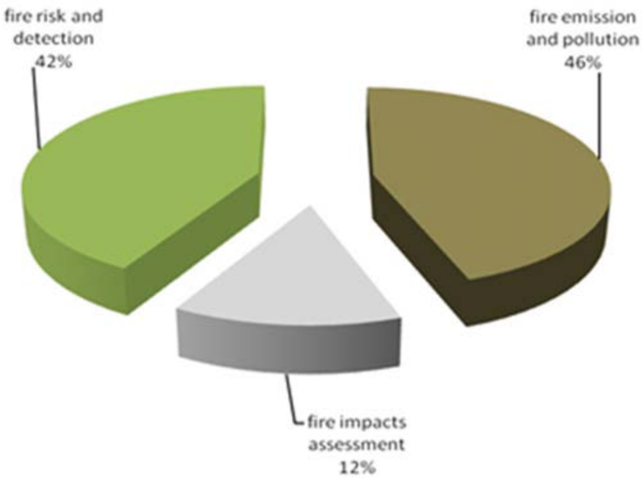

Fig. 5. Cluster 03 of forest and land fire satellite-based research in Indonesia

Research on fire emission and pollution was started in 2002 that elaborate carbon emission from peat and forest fire in Indonesia during 1997 [10] known as the most devastated fire impact in 20th century. Interest on the topic has tremendously increased since 2012 of which smoke/haze issues was included in the research topic [11]. Smoke modelling research seems to become interesting topic for studies and increased in 2015 [12, 13]. The nature of satellite-based data with global accessibility provide appropriate information for fire-smoke relationship and for fire risk and detection and fire impacts assessment researches as well. Similar trend has been observed for fire risk and detection research, whereby hotspot as indicator was comprehensively analysed for fire detection information system [14, 15] as well as fire risk mapping [16, 17]. However, research on fire impacts assessments has been developed earlier back to 2000 [18, 19] and in decreasing trend until 2014 [20].

As for the implication of the information provided by this study, it seems that several aspects of forest and land fire have been studies intensively. However, there are still needs to elaborate more aspects if forest and land fire control in Indonesia is a must. Table 2. Indicates the summary challenges for future fire satellite-based

Table 2. Summary of challenges in fire-satellite-based research in Indonesia

Fire control	Fire aspect	Challenges
Prevention	- Early warning system (EWS) - Fire risk map	- Effectiveness of EWS - fire prone area mapping (national and site level)
Suppression	- Early detection	- user friendly information system (site level) - hotspot accuracy
Post fire rehabilitation	- Fire impact assessments	- research on other impacts (biodiversity) - research on emission factor

5. Conclusion

Satellite-based research on forest and land fire has been developed and enhanced. Undoubtedly, Satellite-based application program plays important role in forest fire control, though, information on the extent and magnitude of the application is still limited; Cluster 03 is most likely secondary cluster for fire-satellite-based research which include: fire emission and pollution, fire risk and detection, and fire impacts assessment; Challenges in the research topics for the future: effective EWS, fire prone area mapping, information system-hostpot accuracy, fire impacts on biodiversity, emission factor; Combining satellite-based and ground researches would achieve optimum results and play important role in minimizing forest and land fire occurrences as well as their impacts in Indonesia.

References

1. Miettinen J, C. Shi, SC Liew. 2011. Influence of peatland and land cover distribution on fire regimes in insular Southeast Asia. Reg Environ Change 2011;**11**:191–201.
2. Boschetti L, D. Roy, and A.A. Hoffmann. Collection 5 MODIS Burned Area product Collection 5 MODIS Burned Area product User's GuideVersion 2.0, November 2009.
3. Chuvieco E, R G Congalton. Application of Remote Sensing and Geographic Information Systems to Forest Fire Hazard Mapping. REMOTE SENS. ENVIRON.1989;**29**:147-159.
4. Feldman R, Sanger J. The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data. Cambridge, UK: Cambridge University Press; 2007.
5. Agrawal R and Srikant R, "Fast algorithms for mining association rules," Proceedings of 20th Int. Conf. Very Large Data Bases, VLDB, Eds. Morgan Kaufmann, 1994: 487-499.
6. Han J, Kamber M, and Pei J. Data Mining: Concepts and Techniques. Massachusetts (US): Morgan Kaufmann; 2012.
7. Guandong X, Yanchun Z, Lin L. Web Mining and Social Networking: Techniques and Applications. New York (US): Springer Science & Business Media; 2010.
8. Syaufina L. 2015. Research status on the relationship between forest fire and biodiversity and human dimension in Indonesia. A Paper presented in the National Seminar on Promoting Research on Forest and Land Fires Mitigation, Adaptation and Impact to Human and Biodiversity, SEAMEO BIOTROP, Bogor, Indonesia, 28 April 2015.
9. Syaufina L. Status of Peatland Fire Research in Indonesia. A paper presented in The 1st International Seminar on Tropical Silviculture, 21 August 2015, Bogor, Indonesia.
10. Page SE, F Siegert, JO Rieley,H-DV. Boehm, A Jayak, and S Limin. The amount of carbon released from peat and forest fires in Indonesia during NATURE VOL 420 7 NOVEMBER 2002. 1997;**420.**
11. Zender C S, AG Krolewski, MG Tosca, and JT Randerson.Tropical biomass burning smoke plume size, shape, reflectance, and age based on 2001–2009 MISR imagery of Borneo. Atmos. Chem. Phys. 2012;**12**:3437–3454.
12. Heriyanto, E., L. Syaufina, S. Effendy. 2015. Forecasting simulation of smoke dispersion from forest and land fires in Indonesia. Procedia Environmental Sciences 2015;**24**:111-119.
13. Nuryanto DE. Simulation of forest fires smoke using WRF-Chem model with FINN fire emissions in Sumatera Procedia Environmental Sciences 2015;**24**:65 – 69.
14. Nurpratami ID, IS Sitanggang. Classification Rules for Hotspot Occurrences Using Spatial Entropy-based Decision Tree Algorithm. Procedia Environmental Sciences 2015;**25**:120-126.

15. Usman, M., IS Sitanggang, L Syaufina. Hotspot distribution analyses based on peat characteristics using density-based spatial clustering. Procedia Environmental Sciences 2015;**24**:132-140.
16. Samsuri, I Nengah Surati Jaya, Lailan Syaufina; Model Spasial Tingkat Kerawanan Kebakaran Hutan dan Lahan (Studi Kasus Propinsi Kalimantan Tengah); Foresta Indonesian Journal of Forestry Faperta USU - Persatuan Peneliti Kehutanan Sumut; No.ISSN: 2089-9890; Maret; 2012;**1**(1):12-18.
17. Sugiarto, D. P., K. Gandasasmita, dan L. Syaufina. 2013. Analisis Risiko Kebakaran Hutan dan Lahan di Taman Nasional Rawa Aopa Watumohai dengan Pemanfaatan Pemodelan Spasial.Majalah Ilmiah Globë. Volume 15 No 1 Juni 2013, 2013;**15**(1): 68—76.
18. Siegert F and AA Hoffmann. The 1998 Forest Fires in East Kalimantan (Indonesia): A Quantitative Evaluation Using High Resolution, Multitemporal ERS-2 SAR Images and NOAA-AVHRR Hotspot Data. REMOTE SENS. ENVIRON. 2000;**72**:64–77.
19. Ruecker G, F Siegert. Burn scar mapping and fire damage assessment using ERS-2 Images in East Kalimantan, Indonesia**.** International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII, Part B7. Amsterdam 2000:**33**
20. Nagai S., R Ishii, S Affendi, H. Kobayashi, M Matsuoka, T Ichie, T. Motohka, JJ Kendawang, R. Suzuki. International Journal of Remote Sensing. 2014;35(23):7910-7926.