Prediction of Soil Erosion Depth Due to Increase in Forest fire Danger Rate by Data Mining

Divya T.L
Assistant Professor,
Dept of M.C.A
R.V.College of Engineering,
Bangalore, India.

Dr. Vijayalakshmi M.N
Associate Professor,
Dept. of MCA,
R.V.College of Engineering,
Bangalore, India.

ABSTRACT
Forest areas and unfarmed land fires have been witnessed in the history for loss of soil nutrient. Forest Fires is one of the sources of trouble for a long time. Fires for larger hectare have huge pressure over the ecological system [2] .It is required to calculate the fire danger rate, loss of land area due to forest fire, which further helps forest management in identification for loss of wild life in particular area. General Unary Hypotheses Automaton **(**GUHA) is used to predict the forest land area burnt in hectare [4].The fire danger rate is identified from the obtained result. Rules are generated on meteorological conditions. The patterns are recognized using temperature, rainfall and wind speed. This paper intends to identify the item set for loss of soil depth due to forest fire. Apriori algorithm is used to obtain frequent item sets. The item set are classified into rules for result.

Keywords
General Unary Hypotheses Automaton**,** Apriori algorithm, Forest fire, patterns.

1. INTRODUCTION
Forest fires caused the relocation of susceptible populations as well substantial damages amounting to millions of rupees. Fires caused huge damage in the year 2007 affecting huge territories besides notable number of human sufferers [2]. To predict the fire danger rate , forest burnt area is used. Using General Unary Hypotheses Automaton forest burnt area is identified [6]. From the result fire rate is obtained to predict depth of soil loss.

2. SOIL EROSION
Soil naturally loses the nutrients by the act of water or wind and fire. The potential for severe soil erosion is a consequence of forest fire. When fire occurs it burns plant material. Forest fire are major concern throughout world, soil erosion is one of problem .Since total afforestration is virtually for reforestation, efficient planning of resource allocation is required. 19.27% or 63.3 million hectare of the Indian land hectare has been classified into forest area as per the report of the Forest Survey of India [2], of which 38 million hectare alone are accumulated with resources in great quantity. Thus forests look enormous difficulty in soil erosion and loss of wild life. Indian forests are also in risk due to the forest fires primary to their degradation [2].

3. OVERVIEW OF ALGORITHMS
Algorithm maps data into predefined groups or classes. It is referred as supervised learning because classes are determined before examining the data [3].Classification algorithms require that the classes be defined based on data attribute values .They look for characteristics of data which is already known [3].

3.1 General Unary Hypothesis Automaton (GUHA)
General Unary Hypotheses Automaton is a method of automatic creation of hypotheses based on experimental records, hence a means of data mining. GUHA procedure will collect the data. It provides simple definition of a large set of applicable patterns and at last generates and verifies the applicable patterns. Finally gives the prime data [6].

The association rules have the form $\Phi \approx \psi$ and it is possible to mine for conditional association rules in the form $\Phi \approx \psi / \chi$ as well. Here Φ, ψ and χ are conjunctions of Boolean attributes automatically derived from many-valued attributes in different ways of antecedent, succedent and condition respectively. The symbol $\approx$ is called a 4ft-quantifier. The association rule $\Phi \approx \psi$ means that Boolean attributes Φ and ψ are associated in the sense of the 4ft-quantifier [4].

3.2 Apriori Algorithm
The Apriori is well known association rule algorithm which is used to generate candidate item sets of a particular size. Next Generated item set will be compared with larger item set. Apriori classification algorithm is used to classify the meteorological data. Generating association rules is straight forward. Association rule π is implication of the form A=>B, where A is Confidence and B is consequence of different set of attributes [8].

4. IMPLEMENTATION
The vegetation considered is dry forest area. Soil considered for nearly 15cm from the land surface with three categories. They are surface level, superficial, subterranean. The meteorological data considered here are Temperature T, Rainfall RF and Wind speed WS. Here generated rule has to satisfy all conjunction to give $\Phi \approx \psi$ Antecedent $\Rightarrow$R, WS and Succedent RF [4].
Antecedent: conjunctions of Temperature T and Wind Speed WS. Succedent: conjunctions with Rain fall RF [4].
Condition: (Temperature T>30 & Rainfall RF<1.0 & Wind Speed WS>28) =1.

Temperature (T) is transformed with four different categories. It measures temperature by Celsius. Temperature T between 0 to 20 is considered as Cool C, T between 20 to 30 will be moderate temperature, T between 30 to 40 celsius will be high and T more than 40 celsius will be very high temperature [4].
The values of Rainfall (RF) is measured in mm/hr and transformed into three categories. If rainfall RF is less than1.0mm/hr it is low and RF between 1 to 15mm/hr is high and more than 15mm/hr will be very high[4].
The Wind Speed WS is measured in km/hr and transformed into five categories. WS is less than 1km/hr means peace full wind, Wind speed WS between 1 to 28km/hr means fresh wind, WS between 28 to 75km/hr means strong wind, WS between 75 to 115 km/hr means storm and more than 115km/hr means hurricane [4].
Rules: (T>30 & RF<1.0 &WS>28) =1 means more than 100ha land will affected by fire in the forest. The 4ft-quantifier is formally a boolean condition concerning the four-fold contingency table with frequencies w, x, y, and z[4].
The 4ft-quantifier is given as

w	x	r
y	z	s
k	l	m

Where w=A (Φ≈ψ) the number of objects in the data satisfying both Φ and ψ; *x*=A(Φ & ¬ψ) (satisfying Φ but not satisfying ψ), *y*=A(¬Φ & ψ), z=A(¬Φ & ¬ψ); *r*, *s*, *k*, *l* are minor, *r=w+x*, *r=y+z k=w+y l=x+z* and in is the cardinality of the set of objects (the number of rows of the data matrix, *m=w+x+y+z*). Association means, there are enough coincidences and not too many differences k, l, r and s. Thus a quantifier *q*(w,x,y,z) is associational if *q*(w,x,y,z)=1 and imply *q(w',x',y',z')*=1. Φ Giving Temperature and Wind Speed (T& WS) and ψ giving Rain fall RF

Table.1 Rule Generation

Sl. No	T	RF	WS	Boolean Value	Wild conservation area affected due to fire(A)
1	C	V	P	0	No Occurrence of fire
2	M	L	SW	1	Less Area (<100ha)
3	H	L	S	1	More Area(>100ha)
4	VH	L	H	1	Greater Area(>500ha)

GUHA prediction rules for the patterns have given loss of land area during forest fire. Temperature is cool, rainfall is very high and low wind speed, fire will not occur. If temperature is medium, low rainfall and presence of Strong wind will cause fire in 100ha of land. Suppose temperature is high, rain fall is low and wind speed is more than 75 km/hr it will cause fire in more than 100ha. Temperature is very high, rainfall is also low, Wind speed is more than 115km/hr, it will cause fire in more than 500ha [4].From the result till 30degree centigrade at in dry forest area fire will not occur. This shows no loss of soil nutrients. Suppose 5000ha dry forest area is having temperature T >30centigrade there is chance to occur fire.

Boolean value 0 shows no occurrence of fire the fire danger rate is 0. Fire danger rate in dry forest of 5000ha in temperature T>30 centigrade, Strong wind speed and Low rain fall is given as less area of 100ha.Temperature is high, rain fall is low and wind speed is more than 75 km/hr it will cause fire in more than 100ha.and less than 500ha.Temperature is very high, rainfall is also low, Wind speed is more than 115km/hr, it will cause fire in more than 500ha.10 % is the highest value obtained. Threshold support can vary between 0 to 9.98%. The Item set can be given in table 2.

Table.2 Fire danger Rate from GUHA Rule Generation

Sl. No	Boolean value from GUHA (Table1)	Fire Danger rate (%)	Description
1	0	0	No fire
2	1	1-1.98	Forest Fire with in100ha
3	1	2-9.98	Forest fire with in 500ha
4	1	10	Forest fire more than 500ha

Soil depth is classified into ten different levels .Surface level is considered for 0cm to 5cm, superficial is considered from 6cm to 14 cm and Subterranean is considered for depth of 15cm.

Table.3: The Soil depth in Dry Forest Area

Sl. No	Soil Erosion depth(cm)	Description
1	0	Dry Land Surface level
2	1	1 cm depth from land surface level
3	2	2 cm depth from land surface level
4	3	3 cm depth from land surface level
5	4	4 cm depth from land surface

		level
6	4.25	4.25cm depth from land surface
7	4.5	4.5 cm depth from land surface level
8	5	5 cm depth from land surface level
9	5>soil<15	Superficial depth from land surface
10	Soil>=15	Subterranean highest depth from land surface

Efficiency of Apriori can be improved by Hash based Technique [8]. 2 item set whose corresponding count in hash table is below the support threshold cannot be frequent. Hence it can be removed from candidate set [8]. Rule generated for soil erosion depth with fire rate is given in table 3

Table.3 Rule Generation for Soil erosion depth from Improved Apriori

Sl. No	Fire danger rate(%)	Soil Erosion depth(cm)	Prediction rate (%)
1	0	0	0
2	1.98	3	19.8
3	2	5	20
4	9.98	14	99.8
5	10	15	100

Efficiency of Apriori algorithm can be improved from hash based technique. Using hash table the result obtained is shown in the table 3.The result of adding hash based technique has filtered the item set whose corresponding count in hash table is below the support threshold. If fire rate is 0 no erosion takes from fire, when fire rate is 1.98%, soil erosion 0takes for 3cm. If fire danger rate is between 2% to 9.98% soil erosion takes from 5cm to 14cm. 10% highest threshold up to 15cm of depth will lose the soil nutrients. The result is shown in the graph.

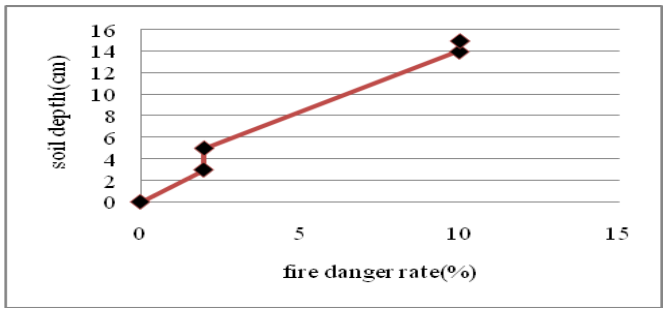

Fig1: Variation in soil depth due to fire danger rate.

Graph represents the cause of soil erosion due to fire danger rate for improved Apriori algorithm. 0 means no Occurrence of fire, No soil erosion. 1.98 fire rate means Soil nutrients loss will takes place for 3 cm to 4cm depth. At the fire rate 2% to 9% Soil nutrients loss will takes place for 5 cm depth. 9.98 fire rate shows Soil nutrients loss will takes place for 14 cm depth. 10 fire rate shows Soil nutrients loss will takes place for more than 15 cm depth.

5. CONCLUSION

By knowing the behavior of fire we can easily predict soil erosion depth in wild conservation area. This prediction also helps for forest management department in reconversion of forest. The forest management can also use this information in further identification for loss of wild life in forest burnt area. This will also help in investigating the rules for optimizing soil erosion.

6. REFERENCES

[1] Ku Ruhana Ku-Mahamud,Khor Jia Yun ,Pattern Extraction and Rule Generation of Forest Fire Using Sliding Window Techique,international journal of Computer and Information Science,Vol.2,No.3,Aug-2009.

[2] K.Angayarkkani, Dr.N.Radhakrishnan An Intelligent System For Effective Forest Fire Detection Using Spatial Data ,(IJCSIS) International Journal of Computer Science and Information Security,Vol. 7, No. 1, 2010.

[3] Data Mining Introductory and Advanced Topics by Margaret H.Dunham, Association Rules pg no178-205, and Spatial mining pg no 235-255, low price edition, Pearson Education

[4] Divya TL,DrVijayalakshmi MN Wild Conservation Burnt Area Prediction By Datamining,2nd National Conference on Information & Communication Technology[NCICT-11],23-24 Dec 2011,pp15.

[5] Aggraval, R. et al.: *Fast Discovery of Association Rules*. In Fayyad, U.M. et al.: Advances inKnowledge Discovery and Data Mining, pp. 307-328, AAAI Press / MIT Press, 1996

[6] Relational Data Mining and GUHA, Tom Karban, K. Richta, V. Sn´aˇsel, J. Pokorn´y (Eds.): Dateso 2005, pp. 103–112, ISBN 80-01-03204-3.

[7] S.Nandagopal, S.Karthik, V.P.Arunachalam,"Mining of Meteorological Data Using Modified Apriori Algorithm", European Journal of Scientific Research, ISSN 1450-216X Vol.47 No.2 (2010), pp.295-308.

[8] Data Mining Concepts and Techniques By Jiawei Han and Micheline Kamber, Mining Frequent Patterns, Associations and Correlations Pg227-272, Second Edition, ELSEVIER.