

A Modified Redescriptive Data Mining Technique to Relate Cyclones and Temperatures of Coastal AP, India

M. Kamala Kumari
Department of CS

AKNU, Rajahmundry, India.

Prof. P. Suresh Varma
Department of CS

AKNU, Rajahmundry, India

ABSTRACT

Redescription Mining, not a new problem describes set of objects or entities in at least two ways either with Boolean, Categorical or Real data. With the consideration of real valued data, the constraints that are applied with some range of values given more scope to the accuracy of redescrptions and that are to be statistically significant.

Considering Geographical area, pertained to some specific regions of Andhra Pradesh, India, in this paper we consider the Natural Hazards and the effect of Temperature on those areas that can describe the regions in two ways, either by the variations in temperature or by the effect and severity of natural hazards. An attempt is made in this paper, to show the correlation between frequency of Natural hazards and temperature increase in the coastal areas of Andhra Pradesh. The algorithm proposed works for real-data on either side of the description. Nine coastal districts of Andhra Pradesh are considered. Study on these districts shows how these districts are more vulnerable to Cyclones in specific, when compared to other districts. Experiments on this resulted in the redescription of coastal areas with temperature and cyclones.

Keywords

Redescription Mining, Natural Hazards, Coastal Andhra Pradesh, Data Mining.

1. INTRODUCTION

Whatever the domain we focus, there exists huge data associated with it. Mining is the activity of extracting the information hidden in the large database, but there also exists redundancies in describing that data. So we need to find that redundancy of description and if we identify at least two different types of description for the same set of entities then we call that as re-description. Finding multiple ways to characterize the same entities is a problem in today's science domains. In Science, related to Medicine, D. Kumar³[2007] said for example, we may typically want to find subset of patients with a combination of similar symptoms and related genes. As another example, according to Redescription Mining research of E.Galbrun⁴[2011], the bioclimatic constraint that must be met for a certain species to survive constitutes that species bioclimatic envelope.

In Geological Science, The Natural Hazards, NH are directly proportional to the temperature effect for some typical regions. Global warming is an issue for the fluctuations in temperature and NH, M.K.Kumari⁶[2012]. There exists a positive correlation between NHs and temperatures. Our analysis study has shown that there exists a positive correlation between the temperature increase and the

frequency of NH. We try to figure out the specifications in the temperatures and NHs by considering the real-valued data on both sides. Say for example, we can say that a specific area is re-described by stating that the area of South coastal region of Andhra Pradesh having Cyclones of winds ranging from 47m/s to 50m/s wind, is also the area with the average mean temperature of 42⁰ Celsius to 47⁰ Celsius during the months of May and June.

In our re-description mining, we consider the input containing regions described by two sets of characterizing variables. One set of variables contains Natural Hazards and the other set of variables contains temperature spread across the years. The task is to find a pair of queries which returns the same subset of entities where the queries contains the combination of variables of temperature and Natural Hazards and entities are nothing but the regions considered. We see that the set of entities are the same in both the domains.

Motivation:

Until now re-description mining algorithms are well implemented with Boolean data. Esther⁵[2011] Proposed Reremi algorithm which considers real-valued data. We have considered real valued data of temperature and intensity of Natural hazards. The constraints, Minimum threshold value for Jaccard Coefficient which measure similarity, independency of two random queries, number of entities in a redescription, Number of entities contributed by a variable, Number of variables in a query, Statistical significant of the queries and the type of the query are considered in this paper.

Andhra Pradesh and its Coastal Districts.

Andhra Pradesh is the 4th largest state in India by an area of 2,75,608 km² and 5th largest population spread over 23 districts. Although it lies in the central part of the Peninsular Indian shield it has prone to many disasters. The state has a coastline of approximately 1030 km and an equal history of major Natural hazards like Cyclones, Earthquakes and Tsunamis. Coastline has a total of 9 districts and among them 2482 are villages.

Why we should relate Natural Hazards and Temperature:

According to NASA and NDA, the primary consequences of Global warming are the rise in temperature and rise in sea level. It is predicted that a rise of as little as temperature 40⁰C could lead to the melting of glaciers, ice sheets, ice caps around the world. Melting of Glaciers will lead to Earthquakes under certain specific pressures, M. Usman⁸ [2011]. An increase in temperature also could lead to more frequent heat waves, higher rates of evaporation and precipitation, which could certainly increase the strength and frequency of tropical cyclones, floods and droughts.

Fig 1 shows the occurrences of Natural hazards illustrating the consequences of Temperature rise. Changes in climate affect not only average temperatures but also extreme temperature increases the likelihood of generating Natural Disasters.

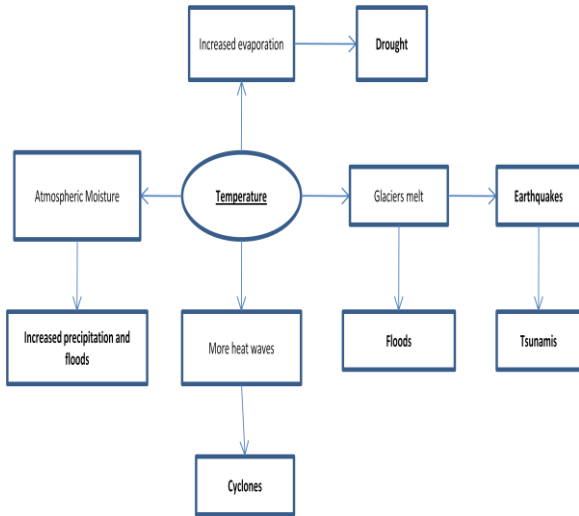


Fig 1: Occurrences of hazards due to temperature

Climate change is predicted to have a range of serious consequences, some of which will have impact over the longer term, like spread of disease and sea level rise, while some have immediately obvious impacts, such as intense rain and flooding. While recognizing the importance of the other predicted consequences of Natural hazards, this paper focuses on identifying those areas which are affected by them due to rise of temperatures.

2. CYCLONES IN ANDHRA PRADESH

Andhra Pradesh is battered by every kind of natural disasters: Cyclones, Floods, Earthquakes and Droughts. The coastal region suffers repeated cyclones and floods. The 1977 cyclone and tidal wave, which resulted in great loss of life, attracted the attention of the central and state Governments of India and the international donor communities, as did those of 1979, 1990 and 1996. The floods in the Godavari and Krishna Rivers caused havoc in the East and West Godavari and Krishna districts. More than seventy cyclones have affected AP last century. According to Disaster Management Department, DMS Government of Andhra Pradesh, the history of 118 years Cyclones data from 1892 to 2010, shows 72 Cyclones which crossed Andhra Pradesh Coast have affected only all nine coastal districts. The incidence of cyclones seems to have increased in the past decades, to the extent that severe cyclones have become a common event occurring every two to three years, repeatedly and severely affecting the state's economy while challenging its financial and institutional resources. Almost 2.9 million people are vulnerable to cyclones and their effects in Coastal AP, 3.3 million of who belong to communities located within five km of the seashore. Recently, [11], Cyclone 'Laila' in May 2010 is at an extreme with a death toll of 50 from coastal districts. Fig 2 (Obtained from The Hindu News paper dated: 18th May 2010) shows the 'Laila' affected coastal districts of Andhra Pradesh. Very recently from October 28th 2012 to 31st Cyclone Nilam, the deadliest Cyclone since 'Jal' in 2010 has affected more than 75 deaths (Only to coastal districts of AP). Wind speed of this is 100 km/hour. Affected districts of

Nilam are Nellore, Chittoor, Prakasam, Ongole, East Godavari and Visakhapatnam.



Fig 2: 'Laila' effected areas in Coastal AP, India

2.2 Effect of Temperature on Cyclones in Coastal Andhra Pradesh

Andhra Pradesh is located at 12°41' and 22°N latitude and 77° and 84°40'E longitude. Summers in Andhra Pradesh lasts from the month of March to June. A comparative temperatures with Rayalaseema, Telangana and Coastal districts shows coastal districts record more temperatures in May and June. Temperature generally ranges from 30° to 45°C. Due to these higher temperatures, water at sea level increases. And hence, the higher sea temperature is the higher sea level. This sea level rise causes inundation of coastal habitats for humans as well as plants and animals and more powerful Cyclone surges that can devastate low-lying areas. According to weather experts, higher ocean temperatures cause more frequent and stronger Cyclones. United States Environmental Protection Agency, EPA states that, The Green house gasses stay in the atmosphere for hundreds of years and even if we stabilize this emission, the surface air temperature continue to warm. This is because the oceans which store the heat takes many decades to fully respond to higher green gas concentrations. The oceans response to these concentrations and higher temperature will continue to impact climate over the next several decades to hundreds of years. So we can derive the fact that the major temperature to an area especially the coastal one leads to climatic effects like Cyclones, Storms etc., Fig 3 illustrates the said derivation.

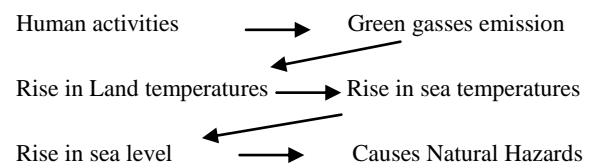


Fig 3: Cause for Natural Hazards

With a frequency of four cyclones per year, one of which usually becomes severe, the Bay of Bengal accounts for seven percent of the annual tropical cyclone activity worldwide. Once the cyclones enter the mainland, they give way to heavy rains which often translate into floods, as it was the case with the damaging cyclone-induced floods in the Godavari delta, in August of 1986. Coastal AP is spread from 13°N to 19°N latitude and from 79°E to 85°E longitude. In tropical regions

the places situated nearly at sea level record relatively large amount of variation in the daily pressure. Andhra Pradesh is the second largest state to be effected by the Cyclones. Andhra Pradesh has the longest coastline of all the states in the country. The 760km length along the sea has laid bare the state to the fury of cyclones that have been a regular feature in the Bay of Bengal. Fig 4 shows the number of Cyclones in

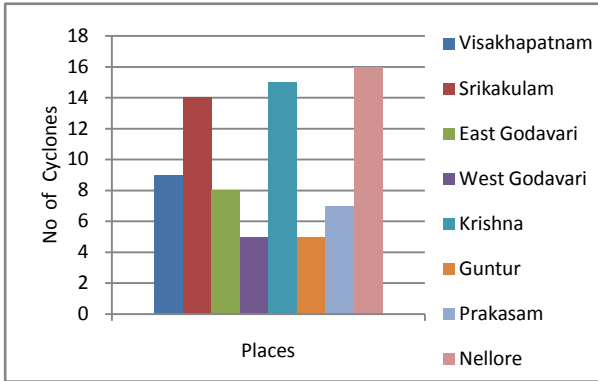


Fig 4 : Number of Cyclones in the Coastal districts of AP

Coastal districts of AP according to the data given by IMD. May month temperature variations, is shown in Fig 5 for the same districts during the same period. The data is obtained from Water Portal India and from that only May mean temperatures of the years from 1995 to 2002 is considered. There is significant direct relation between the effect of Temperatures and the origination of Cyclones on the Coastal Districts of Andhra Pradesh for the considered time period.

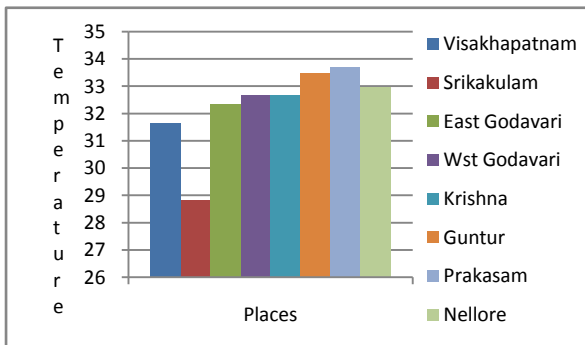


Fig 5: Average Temperatures for the May month

3. OUR APPROACH

Redescription Mining has been extended to consider categorical and real valued data. An algorithm has been constructed which computes the optimal discretization dynamically. We present experimental studies with synthetic data to verify that our algorithm returns good results. We also assess the significance of the results by testing them against different null models. Our Primary application of real valued redescription mining is identifying the regions in South Coastal Andhra Pradesh which are prone to Cyclones as well as heavy temperatures.

3.1 Notations and Definitions

For any object to describe, consider four different types of attributes. Boolean attributes, Real-valued attributes, Ordinal

attributes and Categorical attributes. Consider Boolean attributes and real valued attributes in experimentations. Two sets of attributes A_L and A_R are considered for two different domains. Redescriptors are framed over these two attributes. The set of entities, which are going to be described by A_L and A_R is denoted by E . Represent the data using two matrices D_{CL} and D_{TR} . Both matrices have $|E|$ rows and D_i has $|V_i|$ columns. The value of $D_{CL}(i, j)$ is the value of $a_i \in A_L$ for $e_i \in E$. If I is a set of row indices (or a characterizing vector thereof), $D(i, j)$ is the column j of D restricted to the rows in I . The data is a 5-tuple $D = (A_L, A_R, E, D_{CL}, D_{TR})$. Identify variables in A_L and A_R with the corresponding columns in D_{CL} and D_{TR} when there is no risk of ambiguity.

Consider all four types of variables which are afore mentioned. If $a \in A$ then we interpret the column corresponding it as the truth value assignment for $e \in E$ normally, if it is real-valued and ordinal we consider an interval $[a, b]$ and the truth value assignment induced by the relation $a \in [a, b]$. Special case can be when a is categorical. Then consider $a = c$ for some category of c . These truth value assignments are denoted by notation $[a \leq a \leq b]$ is the Boolean vector that has 1 in the rows where $a \in [a, b]$, and 0 if not. $[a=c]$ is defined analogously. The literals come out from these truth value assignments and their negations for attributes in A . Notice that there are infinitely many intervals yielding the same truth value assignment for some real-valued $a \in A$. To avoid ambiguity, consider only the shortest interval yielding some truth value assignment.

Literals are combined with Boolean operators \wedge (and), \vee (or) and \neg (negation). A Boolean formula is obtained by combining literals with Boolean operators in between. A query over A is a Boolean formula with literals of A . A redescription R of $D = (A_L; A_R; E; D_{CL}; D_{TR})$ is a pair of queries (q_L, q_R) over A_L and A_R , respectively. For a redescription $R = (q_L, q_R)$, where q_L is the boolean formula of left hand side data, i.e of one domain and q_R is the boolean formula of right side data i.e., another domain data. Use $A_L(R)$ to denote the attributes of q_L ; $A_R(R)$ is defined analogously. The support of a redescription $R = (q_L; q_R)$, $supp_D(q_L; q_R)$, is the intersection of supports of q_L and q_R , $supp(q_L, q_R) = supp(q_L) \cap supp(q_R)$. A redescription $R = (q_L, q_R)$ is exact if and only if $supp(q_L) = supp(q_R)$. If a redescription is not exact, it is approximate. The accuracy of a redescription $R = (q_L, q_R)$ is measured using the Jaccard Coefficient, $J(R)$.

$$J(R) = J(q_l, q_r) = \frac{|supp(q_l) \cap supp(q_r)|}{|supp(q_l) \cup supp(q_r)|}$$

Formally, redescription mining is approached as follows: Given data $D = (A_L; A_R; E; D_{CL}; D_{TR})$ and a set of constraints C , and all redescrptions R_1, R_2, \dots of D that satisfy constraints in C . The constraints we took are as follows:

- Accuracy which is based on given threshold value.
- Redescrptions should be statistically significant
- Adding a literal should give any new information
- The type of the query.
- Size of the support of the redescription
- Number of entities by which each variable contributes to it.
- Check the independency of two random queries.

3.2. EREREDM Algorithm

A typical approach to mine such patterns is to first fix the intervals of the attributes we consider and then fix set theoretic expressions and then search in the space of possible instantiations.

Skeleton of the algorithm: We use an approach EReReDM analogous to ReReMi[4], to explore the solution space as well as to improve the efficiency. The algorithm starts redescrptions from singleton redescrptions to redescrptions with logical operators in between until the number of limited attributes in both the queries.

Algorithm EReReDM(D,R)

I/P :

$D = (A_L, A_R, E, D_L, D_R)$

D_L : = Left database, D_R = Right database, A_L = Attributes of D_L , A_R = Attributes of D_R

$r_s = 0$, positive integer to hold Initial Singleton redescrptions,

r_i = positive integer to hold Best Intermediate redescrptions for each iteration

C = Constraints on the Redescrptions to filter and for accuracy of the algorithm.

η_e = Number of entities involved in Redescrptions

R_a = Number of attributes in each Redescription

J_c = Threshold value for Jaccard Coefficient

E = Total number of entities

I = Number to hold intermediate singleton redescrptions

O/P: A set of Redescrptions, \mathcal{R}

$\mathcal{R} \leftarrow \emptyset$; $I = \emptyset$; $m = r_s$

//Find out first best Initial Redescrptions

for each $a_i \in A_L$

for each $a_j \in A_R$

if $\text{supp}(a_i, a_j) > J_c$

then $I \leftarrow I \cup (a_i, a_j)$ and $r_s \leftarrow r_s + 1$

select best r_s best redescrptions

with the highest Jaccard

Coefficient order and leave the

remaining.

$I \leftarrow \{ r_s \text{ best Singleton Redescrptions} \}$

for each of $S \in I$ do

$K \leftarrow \{ S \}$

// get free variables(not yet included in Redescrptions) to add to the obtained RDs

$F_L(S) \leftarrow$ All left side variables in D_L not in S

$F_R(S) \leftarrow$ All right side variables in D_R not in S

If $F_L(S) \neq \emptyset$ and $F_R(S) \neq \emptyset$ then

$\epsilon \leftarrow \{ S \}$ **// Set of RDs to be extended during the**

// next iterations.

while $\epsilon \neq \emptyset$ do

for each $R_i \in \epsilon$ do

for side, $s \in \{ \text{left, right} \}$ and operator $o \in \{ \wedge, \vee \}$

and $l_L \in F_L(S)$ $l_R \in F_R(S)$ do

If R_i can be extended on side s with o and l

$\{ F_R(S), F_L(S) \}$ then

$K \leftarrow K \cup \{ \text{Best such extension of } R_i \text{ admitting}$

constants $C \}$

$K \leftarrow \text{bestof}(K, r_i)$

$\epsilon \leftarrow \{ R_i \in K : F_L(R) \neq \emptyset \text{ and } F_R(R) \neq \emptyset \}$ **// end of while**

loop

$R \leftarrow R \cup K$

// check constraints on R

Let J_c = Given threshold value

Let $P_R = (|\text{supp}(q_L)| |\text{supp}(q_R)|) / |E|^2$

//Check the independency of two random queries, q_L and q_R in R using Poison distribution.

$pvalM(q_L, q_R) \cong \sum_{S=|\text{supp}(q_L, q_R)|}^{|E|} e^{-|E| * P_R} * ((|E| * (P_R))^S) / S!$

If $pvalM(q_L, q_R) > J$ then \mathcal{R} can be discarded

Let N_e = No: of entities in $R(q_L, q_R)$

If $N_e < \eta_e$ then \mathcal{R} can be discarded

Return \mathcal{R}

3.3 Framework of the algorithm

This algorithm is similar to Reremi algorithm[4] with extensions and changes to it. The outline of the algorithm follows the example given above. To find out the boundaries of the solution space we followed beam-search. Beam-search is heuristic search algorithm, that explores a graph by expanding the most promising node in a limited set. Here consider most promising queries starting from singleton queries and then expand them by adding most promising variables and logical operators in between. The queries are constructed bottom-up, starting from single variable on both sides of the queries. This is progressively expanded by appending appropriate operators and attributes. For example we could start with a pair $(c, \neg t)$ as initial redescription and then extend this in turn as $(c \wedge eq, \neg t)$ $(c \vee eq, \neg t)$ $(c \vee \neg eq, \neg t)$ $(c \wedge \neg eq, \neg t)$, etc., After evaluating all possible one step extensions and satisfy Jaccard's threshold, we select the best candidates and extend them in turn.

The EReRedm algorithm stores the two input data of two domains in the form of two matrices, D_L and D_R . Algorithm starts by storing null value in to the Redescription set R . As described above singleton redescrptions are first identified based on threshold value, J_c . We keep r_s the most promising out of them. Let I , hold all these redescrptions to consider one after the other. Considering each one in to S , take F_L as the free variables of D_L except those that are in S . These free variables are added to the existing redescrptions to get full query. Similarly we do for the right hand side query by considering D_R and free variables of it keeping in F_R . As long as these sets are not empty, extend them by side, s either to the left or right and considering operators, o with the attribute, l either from F_L or F_R . This is appended to K . Again ϵ is updated with new Intermediate redescrptions, R_i . This process is repeated until ϵ is equal to empty. Bestof() is a function which will select only r_s best redescrptions based on J_c .

Why Poison Distribution?

Given the support of the queries, whenever we find a Redescription between two queries say, q_l and q_r should carry some new information. To measure this, Esther et al.,[4] had tested a null-model in which the two queries would be independent(For a Redescription, independent queries cannot be a good result). They have computed a p -value that represents the probability that two random queries with two marginal probabilities of them, equal to those of q_l and q_r have an value equal to or larger than $|\text{supp}(q_l, q_r)|$. This

probability they have used binomial distribution. Here we have taken Poisson distribution for probability. This is given by

$$pvalM(qL, qR) = e^{-|E|*Pr} \frac{[|E|*(Pr)]^S}{S!} \quad \text{Eq (1)}$$

where $Pr = |\text{supp}(qL) \cap \text{supp}(qR)| / |E|^2$.

The higher the *p-value* the more the queries are independent and less significant the query is. So for any two queries, qL and qR is *p-value* is larger, then that query combination is not used.

Poisson distribution is widely used discrete probability distribution for the counts of events that occur randomly in a given interval of time or in a given space. We prefer this when N, Number of approaches is very large and probability of each approach will be of infinite. We also use Poisson when the number of situations occur in which we observe the counts of events with in a set unit of time, area, volume, length etc., Eg., Number of cases of a disease in different towns. In such situations we are interested in whether the events occur randomly in time or space or not. The probability of observing *x* events in a given time interval is given by an approximation of Poisson distribution, where $\lambda = n*p$.

$$P(X=x) \approx e^{-\lambda} \frac{\lambda^x}{x!} \quad x=0,1,2,3\dots \quad \text{Eq (2)}$$

Where X is the number of events in a given interval and λ is mean number of events per interval. If n is large and p is small then B(n,p) can be approximated with $P_o(\lambda)$.

If the exact distribution has too much of detail and if there are some features which are irrelevant(The ones which are not satisfying Jc) to the queries we want. So to approximate distribution, we focus on the things that are really concerned with. Also a Poisson random variable can take on any positive

integer value, but in contrast a Binomial Distribution always has a finite upper limit.

Since Natural hazards data is considered with large range of values and each occurrence is of less probability we have chosen Poisson distribution to be the best approximated value for the statistical significance of the query. This we have checked in the algorithm after Redescriptions are found and few of them will be removed which does not satisfy the constraint. Similarly, when appending an attribute, a to a Redescription.

3.4 Checking Constraints on the Redescriptions

For the efficiency improvement of the algorithm, we formulate some constraints and check the compatibility of redescriptions with them. As a preliminary constraint the check on is the accuracy constraint which will be obtained through threshold value for Jaccard coefficient. Basically Jaccard coefficient is used to find the accuracy of similarity between two entities.

4. EXPERIMENTATION

A sample data is taken as temperatures recorded every month for about an year for a particular area. We have the recorded measurements of Natural Hazards recorded in those areas. Consider Earthquakes(EQ), and Cyclones on the left hand side Database and the affected areas of them as shown in Table 1. The measurement for EQs considered are their magnitude and for Cyclones, their Wind speed(km/hr). On the right hand side, we have taken the temperatures for three different months, January, May and October and the affected areas of it. Results are shown in Table 2. Left hand side queries are given on the left side of the table as qL and right handside queries on right side as qR. Three sample redescriptions are derived. Jaccard Coefficient which signifies the strength of each, is calculated. Considerations of these redescriptions depends on the threshold value we regard.

Table 1. Earth quakes, Cyclones and Temperature data

Earth Quakes	Cyclones	Affected Areas	Jan	May	Oct	Affected Areas
2.3	100	Nlr, kavali	23	45	21	Vja, guntur
5.2	121	Ongl	30	43	12	Kvl, ongl
6.3	97.4	Vja, bptl	19	39	11	Kkd
3.2	211.7	Visk,vizngm,kkd	22	45	19	Vsk,vja
1.2	74.8	Gntr, vja	30	34	30	Nlr,gntr, kvl
4.0	234	Ongl,kavali, vja	18	44	17	Kkd, nlr

Table 2. Redescriptions for few areas of Coastal Andhra Pradesh

$1.2 \leq EQ \leq 4.0 \cap 100 < Cycl < 211.7 \rightarrow \{ nlr, kvl, ongl, vsk, vja, kkd \}$	$19 \leq temp(Jan) < 23 \cup temp(May) = 44 \rightarrow \{ kkd, vja, vsk, nlr \}$
Here the four element set, {kkd, vja, vsk, nlr} can be redcribed with either of qL or qR with a strength , JC of $4/6 = 0.667$	
$4.0 > EQ < 6.5 \cup Cyc > 121 \rightarrow \{ ongl, vja, bptl, vsk, viziangrm, kkd, kvl \}$	$30 = temp(Jan) \cup 43 = temp(May) \cap 30 > temp(Oct) \rightarrow \{ nlr, gntr, kvl, ongl \}$
Here the two element set {ongl, kvl} can be redcribed with either qL or qR but with a lesser strength of $2/4 = 0.25$	
$2.0 \geq EQ \leq 3.2 \cap Cyc \geq 100 \rightarrow \{ vsk, viz, kkd, kvl, nlr \}$	$22 > temp(Jan) \cup 21 \geq temp(Oct) \rightarrow \{ vja, gtr, kvl, ongl, nlr, kkd \}$
For this the set {kkd, kvl, nlr } has a redescription with a strength of $3/8 = 0.375$	

5. CONCLUSION

A study has been done on the frequent occurrences of Cyclones and Earthquakes of Coastal districts of Andhra Pradesh, India. Similarly temperature levels are also observed for specific months for a period for the same areas. Causes for the temperature rise and also for Cyclones are mentioned clearly. For the first time redescription of Coastal Andhra Pradesh has been described in terms of the Cyclones, Earthquakes and temperature. A random data has been considered and experimented with the algorithm outlined and results are displayed. The effectiveness of the algorithm has been increased by using Poison Distribution instead of Binomial Distribution in measuring and checking constraints. Much contribution is confined to Cyclones and Temperature. A study on Earthquakes and Tsunamis of Coastal Andhra Pradesh has been done. Challenging task ahead is relating earthquakes and tsunamis with temperature. Earthquakes and Tsunamis which are related should also redescribe the targeted areas. A study on Earthquakes reveals that, it has a direct proportion with temperature below the earth's surface. Temperature at a down-going slab, at a subduction zone when penetrates, rocks under this slab hotter than the incoming temperature are too ductile to generate Earthquakes. Also, due to the severe rise in temperature levels at mountain areas and near north poles, which causes glaciers to melt, will generate earthquakes. On the other side, there is an ambiguity to relate temperature and earthquakes according to few researchers and geologists. A detailed and analyzed study may lead to a corner. Similarly tsunamis are on the same path to relate with temperatures. But earthquakes and tsunamis recount on one side if we see the affected areas.

6. ACKNOWLEDGMENTS

Our thanks to our colleague Dr. Swamy, Geological department, from our University who has contributed geological data for our experimentation. We also thank team of India Water portal for maintaining the monthly wise temperature data. Our sincere thanks to IMD who has

sufficient and necessary data of Cyclones in their Website which has made our work to go insight.

7. REFERENCES

- [1] A. Gallo., P. Miettinen, and H. Mannila. 2008. Finding subgroups having several descriptions: Algorithms for redescription mining. In *SDM*, pages 334–345.
- [2] B. McGuire, April 2010 Potential for a hazardous geospheric response to projected future climate changes Published 19 doi:10.1098/rsta.2010.0080Phil. Trans. R. Soc.
- [3] D. Kumar, 2007. Redescription mining: Algorithms and applications in bioinformatics. Ph.D thesis, Department of Computer Science, Virginia Tech.
- [4] E. Galbrun and P. Miettine , 2011. From Black and White to Full Colour: Extending Redescription Mining Outside the Boolean World. In *SDM*, pages 546–557.
- [5] M.K. Kumari, P. Suresh Varma, Aug 2012. Redescription Mining with Three Primary Data Mining Functionalities. *IJCSIT* vol 4, pages 64–69.
- [6] M.K. Kumari, P. Suresh Varma and A. Praveen., 2012. Redescription Mining with Geological data. In the proceedings of ACC. July 26-28 pp 158-163.
- [7]. Mohammed J. Zaki et al, 2005, "Reasoning about Sets using Redescription Mining" *KDD'05*, August 21–24, Chicago, Illinois, USA.
- [8] M. Usman et al, 2011, Effects of Temperature Increase on Earthquake Frequency and Depth in Northern Pakistan, 2010 International Conference on Biology, Environment and Chemistry, IPCBEE vol 1.
- [9]. Philip N S and Joseph K B, 2001. On the Predictability of Rainfall in Kerala: An Application of ABF Neural Network, In Proceedings of Workshop on Intelligent Systems Design and Applications (ISDA).
- [10] W. Thuiller, B. Lafourcade, R. Engler, and M. B. Araújo. 2009. Biomod – a platform for ensemble forecasting of species distributions. *Ecography*, 32(3):369–373.
- [11] 'The Hindu' News paper dated: 18th May 2010