USE OF SPATIO-TEMPORAL ASSOCIATION RULE MINING FOR HAZARD MITIGATION

ABDUL FATTAH CHANDIO*, AKRAM SHAIKH**, ATTAULLAH KHAWAJA***

ABSTRACT

This paper presents spatio-temporal based association rule mining approach for hazard mitigation. The main aim of this research is to formulate mechanism that can predict the occurrence of hazards. In this paper authors used association rule mining technique for prediction of hazards. In this research history of hazards that collected with reference of time and space is used. The Apriori algorithm based association rule mining is used for prediction of rules that may be used for hazard mitigation. These AI based rules potentially can be helpful to mitigate the catastrophic effects caused by natural hazards.

Key Words: Hazard Mitigation, Association Rule Mining, Spatio- Temporal Data Mining,

1. INTRODUCTION

Hazard mitigation is the most important and most attention required subject of these days, especially in the case of thickly populous countries like Pakistan, China or India. In these countries thousands of people are affected by various types of natural hazards annually. The most of fatalities in Pakistan regarding natural hazards are done because of lack technology and shortage education towards the awareness these hazards. The main theme of this research is the use of spatiotemporal data mining techniques to hazard mitigations and providing the future guidelines of developing hazard mitigation system.

Spatio-temporal applications have been increased in the last decade [1]. In spatio-temporal applications objects are related with each other in complex manner. The spatio-temporal databases provides platform for generation of new set of rules that may encompasses changes [2]. Data mining techniques are typically used for identification of unknown patterns that are embedded inside that data and useful for future strategy.

2. SPATIAL ASSOCIATION RULES

For this research, association rule mining can be used because to detect unknown relationships between different entities. Spatial association rules are those rules showing certain relationships among different geographical and non-geographical attributes which are spatially and temporally associated with each other. These objects/attribute are indicated as predicates [3, 4, 5]. Spatial association rule mining is a main derivative of the spatial data mining [7].

Spatial Association rule mining is of 'IF X THEN Y' format of rules. Here in this research rules are generated in the format of if X then Y followed by (Cover% Conf% Cover Count Sup Count Sup%,) defined in section 6.

3. SPATIO- TEMPORAL HAZARD ASSOCIATION RULE MINING

Here in this research database used for data mining is applied is arranged in a concept of hierarchical classification. Association rule mining of data that arranged in such concept is known as multiple level association rule mining [6,1]. The spatio-temporal process is used to show the response of particular event occurred on a geographical location during certain time interval. Objects can be described by spatial attributes, non-spatial attributes and their relationships with other objects and others indices [8]. It is evident from literature it is found that most of researchers faced problems while applying

*Department of Electronics Engineering QUEST Nawabshah fattahchandio@yahoo.com,
**Department of Computer Sciences MUET Jamshoro akramshaikh@hotmail.com
***Department of Electronics Engineering NED Karachi atta_khawaja@yahoo.com

association rule mining over spatio-temporal data, which contains geographical maps, raster and vector shapes [9]. The use of association rule mining over spatio-temporal data means to dig out certain valuable unknown patterns from spatio-temporal data which are embedded inside. In this research this problem has been addressed by converting geographical data into table format and then processed for association rule mining. In this research the data selected for spatio temporal hazard rule mining is flood hazard data related to Pakistan.

4. FLOOD IN PAKISTAN

Pakistan is one of most severely affected country of flood hazards,139 floods has been recorded in different cities as shown in Figure-1. These floods resulted thousands of life losses and millions of displacements plus billions of revenue losses. The flood hazard study area is comprised of all portions of Pakistan that were actively affected from flood hazards.

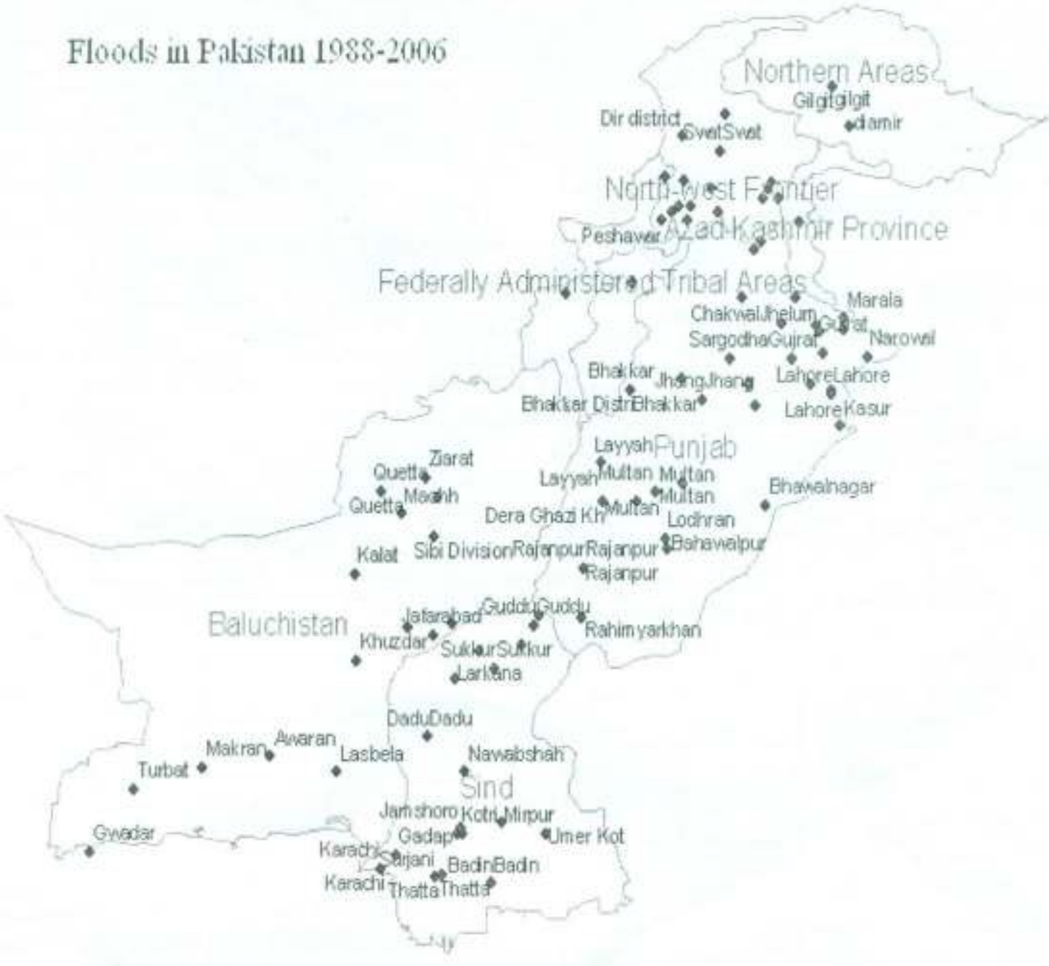

Figure 1: Floods in Pakistan 1988-2006 (name of cities)

The objectives of Association rule mining applied in this research is to dig out the patterns inside hazards database which contains data related to death toll of social population, effected area, names of places and the date of events. Note that the data about casualties were not available specifically so casualty indexes are added fictitiously. Figure-2 Floods in Pakistan 1988-2006 date wise events

Figure 2: Floods in Pakistan 1988-2006 date wise events

As it is mentioned above that it is very much complicated to process association rule mining over spatio-temporal data. Here GIS is used to preprocess spatio-temporal data and convert it into tabular format which can be accessed by association rule mining techniques. For association rule mining the Apriori based data mining software [9] is used in which the spatial data is used in tabular format.

Table 1: History of floods in Pakistan 1988-2006

Name of cities	year	month	Began	Ended	Dead	Displaced	Affected Region (sq km)	latitude	longitude
Bahawalpur	1988	sept	09/21/88	10/08/88	low	152	6,980	29.3900	71.6700
Gujrat	1988	sept	09/21/88	10/08/88	low	122	460	32.5600	74.0600
Lahore	1988	sept	09/21/88	10/08/88	high	2	1,220	31.5710	74.3130
Narowal	1988	sept	09/21/88	10/08/88	medium	3	125	32.1000	74.8830
Sialkot	1988	sept	09/21/88	10/08/88	high	2	1,235	32.5000	74.5170
Wazirabad	1988	sept	09/21/88	10/08/88	low	54	433,500	32.4500	74.1170
Gujranwala	1988	sept	09/21/88	10/08/88	high	123	460	32.1500	74.1800
Multan	1988	sept	09/21/88	10/08/88	high	21	1,430	30.1830	71.4830
Shahdara	1988	sept	09/21/88	10/08/88	none	32	1,232	31.6300	74.3100
Bhakkar District	1990	july	07/10/90	07/12/90	low	21	123	31.6330	71.0670
Charsadda	1991	jun	06/11/91	06/12/91	none	21	212	34.1400	71.7300
Nowshera	1991	jun	06/11/91	06/12/91	high	12	121	34.0170	71.9800
Peshawar	1991	jun	06/11/91	06/12/91	medium	21	1,255	34.0170	71.5500
Karachi	1992	aug	08/10/92	08/15/92	none	21	100	24.8670	67.0500
Bhawalnagar	1993	july	7/8/93	8/13/1993	medium	21	1,021	30.0000	73.2500
Faisalabad	1993	july	7/8/93	8/13/1993	medium	21	1,220	31.4170	73.0800
Gujranwala	1993	july	7/8/93	8/13/1993	high	1	433,500	32.1500	74.1800
Jhang	1993	july	7/8/93	8/13/1993	high	12	22,332	31.5000	72.2200
Kasur	1993	july	7/8/93	8/13/1993	high	4	6,980	31.1170	74.4500
Khanewal	1993	july	7/8/93	8/13/1993	none	12	3,213	30.3000	71.9330
Lahore	1993	july	7/8/93	8/13/1993	high	2323	460	31.5710	74.3130
Lodhran	1993	july	7/8/93	8/13/1993	high	212	122	29.5330	71.6330
Multan	1993	July	7/8/93	8/13/1993	high	122	1,232	30.1830	71.4830
------	------	------	------	------	------	------	----	------	------
------	------	------	------	------	------	------	------	------	------
------	------	------	------	------	------	------	------	------	------
Gilgit	2006	July	28-Jun-05	12-Jul-06	high	10	460	35.9170	74.3000
North Waziristan district.	2006	July	1-Jul-06	13-Jul-06	Medium	50	6,980	33.0000	70.0000

5. RULES AND DISCUSSIONS

Form Table-1 showing the history of floods occurred in Pakistan 1988-2006, association rules has been generated for four main parameters; name of places, year of floods, month of floods and casualties rate which are mentioned periodically in Table-1.

Name of places that was affected by flood:
These attributes contains the names of different places where floods are occurred. Theses attributes has 85 different names of cities.

Year of flood: This shows years of floods that occurred ranging 1988 to 2006

Month of floods: Represents months of floods.

Casualties rate / death rate: This shows the rate of casualties, high, low, medium and none.

Table -2 to Table-5 are showing the statistics of different attributes, after applying association rule mining.

Table 2: Years of floods in Pakistan

Attribute Value	Frequency	Probability
1988	9	6.50%
1990	1	0.70%
1991	3	2.20%
1992	1	0.70%
1993	16	11.50%
1994	6	4.30%
1997	9	6.50%
1998	1	0.70%
1999	5	3.60%
2001	9	6.50%
2002	6	4.30%
2003	36	25.90%
2005	33	23.70%
2006	4	2.90%

Table 3: Months of floods in Pakistan

Attribute Value	Frequency	Probability
Aug	16	11.50%
Feb	8	5.80%
July	89	64.00%
Jun	10	7.20%
March	1	0.70%
May	5	3.60%
Sept	10	7.20%

Table 4: Casualties during floods in Pakistan

Attribute Value	Frequency	Probability
casualities_greater_then_100	51	36.70%
casualities_greater_then_10_less_then100	15	10.80%
casualities_less_then10	34	24.50%
none	32	23.00%

Table 5: Names of places and frequency of floods in Pakistan

Attribute Value	Frequency	Probability
Gwadar	1	0.70%
Awaran	1	0.70%
Badin	2	1.40%
Bagh district	1	0.70%
Bahawalpur	1	0.70%
Bhawalnagar	1	0.70%
Charsadda	2	1.40%
Chiniot	1	0.70%
Chitral	1	0.70%
Dadu	2	1.40%
Dera Ghazi Khan	3	2.20%
------	------------	----------
------	------------	----------
Umer Kot	1	0.70%

In this research 154 rules were generated from association rule mining related to flood disasters in Pakistan. However these rules are interesting and potentially helpful for hazard mitigations. Formation of rules is elaborated in section-6 Rule Format.

6. RULE FORMAT

Each rule has 'If X then Y' sentence followed by five range values parenthesized by square brackets; *(Cover% Conf% CoverCount SupCount Sup%,).* The first one "*Cover%*"indicates the covering percentage, the percentage of occurrence of that attributed event (**X**) in total number of events. The second one *Conf%* is the confidence percentage, this shows percentage of occurrence of second attributed event (**Y**) while first event (**X**) is occurred. The third one "*CoverCount*" is derived by mean of count of X events out of all (flood) events. The fourth one "*SupCount*" is the counting of occurrence of both events (X & Y). and fifth one is the "*Sup%*" percentage of supporting count out of total number of events Here are few temporal rules that were generated.

Rule 1:

class = casualities_greater_then_100
-> month = July
(36.691% 64.71% 51 33 23.741%)

Rule 2:

month = July
-> class = casualities_greater_then_100
(64.029% 37.08% 89 33 23.741%)

Explanation of Rule #1: It shows the casualties more then 100 if flooding month is July, where:

1) **Cover%** =

$$\frac{\text{Total count of casualties greater then 100}}{\text{Total count of floods 1988 - 2006}} * 100 = \frac{51}{139} * 100$$

= 36.691%

2) **Conf%**=

$$\frac{\text{Count of floods with casualties greater then 100 in the month of july}}{\text{Count of floods with casualties greater then 100 out of total flood event}} * 100$$

$$= \frac{33}{51} * 100 = 64.71$$

3) **Total count of floods with casualties greater then 100** = 51

4) **Count of floods with casualties greater then 100 in the month of july** = 33

5) **Sup%** =

$$\frac{\text{Count of floods with casualties greater then 100 in the month of july}}{\text{Total count of floods 1988 - 2006}} * 100 = \frac{33}{139} * 100 = 23.741\%;$$

Similarly we can understand rule #2.

7. CONCLUSIONS

This research shows the importance spatio temporal data mining for mitigation of hazards. The main focus of this research is development of modular strategy based on spatio temporal hazard data mining for mitigation of hazards in Pakistan. This research shows how to identify proper time, that has more probability of floods in Pakistan. After implementation and analyzing association rule mining it is evident from the results that is month of July has more probability of floods in Pakistan. However, identification of useful patterns is complicated task and need sufficient amount of knowledge over spatio-temporal database. In this research, process of spatio-temporal association rule mining for hazard mitigation guidelines is provided. That composed of development of hazard database to implementation of association rule mining for mining out the useful patterns which are potentially helpful for future mitigation.

ACKNOWLEDGEMENT

Authors are willing to acknowledge this research in the name of professors and teaching fellows at, Beijing Institute of Technology Beijing China, whose enormous guidelines, enabled authors to carryout this research.

REFERENCES

[1] Shahram Ghandeharizadeh, "Real-time Spatio-Temporal Databases in Support of Human Motor Skills, In Towards Replacement Parts for the Brain", MIT Press, Editors: Berger, Glanzman, 2003.

[2] Natalia Andrienko, Gennady Andrienko, "Data and task Characteristics in design of Spatio-Temporal Data Visualization Tools", symposium of Geospatial Theory. Processing and Application, Ottawa 2002

[3] Abdulvahit Torun, Sebnem Düzgün," Using Spatial Data Mining Techniques to reveal vulnerability of people and places due to oil transportation and accidents. A case study of Istanbul Strait", ISPRS Technical Commission II Symposium, Vienna, 12 – 14 July 2006

[4] Karasova V, "Spatial data mining as a tool for improving geographical models", Masters Thesis, Helsinki University of Technology, May, 2005

[5] Mohammed J. Zaki, "Scalable Algorithms for Association Mining, IEEE Transactions on Knowledge and Data Engineering", v.12 n.3, p.372-390, May 2000

[6] Kurt Mehlhorn, Mark Ziegelmann, "Resource Constrained Shortest Paths", Algorithms ESA 2000, ISSN 0302-9743 (Print) 1611-3349, Volume 1879/2000, pp. 326–337, January 01, 2000

[7] Lizhen Wang, Kunqing Xie, Tao Chen et al., "Efficient discovery of multilevel spatial association rules using partitions," Information and Software Technology, vol. 47 (13), pp. 829 – 840, 2005.

[8] Zhang Xuewu, SU Fenzhen, D U Yunyan, Shi Yishao, "Association Rule Mining on Spatio-temporal Processes", Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08, IEEE, Oct. 2008

[9] Jerey Mennis "mining Association Rules in Spatio-Temporal Data: An Analysis of Urban Socioeconomics and Land Cover Changes". Transactions in GIS January 2005