

Disaster Impact Mitigation using KDD and Support Vector Machine algorithms

D. John Aravindhar^{1*}, M.N. Sushmitha² and K.Padmaveni³

¹Professor, Hindustan Institute Of Technology and Science, Chennai, India

² & ³Assistant Professor, Hindustan Institute Of Technology and Science, Chennai, India

{jaravindhar,mnsushmitha,kpadmaveni}@hindustanuniv.ac.in

Abstract

Disasters such as Hurricanes, Typhoons, Floods and earthquakes are not good for the society since it causes serious damage for the society. A natural disaster causes loss in property as well as in life of victims. The victims need immediate help once they are affected by the disaster. The immediate needs are rescue, food and communications. The survey says victims of recent disaster were unable to get instant communication regarding the evacuation path and other help from authorities for remedial action. This can be overcome with our proposed idea of having a database of area wise population along with the pre-disaster and post-disaster satellite images of the disaster affected area. Knowledge Discovery in Databases (KDD) is used in data pre-processing and to extract knowledge from the database. Support vector machine (SVM) is used to classify the disaster effect with the pre-disaster and post-disaster satellite images as input. The idea is implemented and tested with sample data and has given impressive results.

Keywords: Knowledge Discovery in Databases, Support vector machine, Satellite

images Received on 09 April 2018, accepted on 03 May 2018, published on 12 June 2018

Copyright © 2018 D. John Aravindhar *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.12-6-2018.154812

*Corresponding author. Email: jaravindhar@hindustanuniv.ac.in

1. Introduction

A natural disaster can cause loss of life as well as property damage. Top ten deadliest natural disasters include Calcutta cyclone in 1737 with death toll of 300000, India Cyclone in 1839 with death toll of 300000 and Indian ocean earth quake & Tsunami in 2004 with overall death toll of 280000. Relief involves moving and keeping emergency equipment; shifting people out of danger; supplying required food, shelter and medical services.

In this paper, we propose an idea of predicting the impact of disaster within a few minutes of its occurrence. This will help in the speedy recovery and help in the evacuation of people to a safer place. The

paper is organized as follows. Section II gives the complete details about review status of research and development in this subject. Section III illustrates the proposed methodology using KDD and SVM. The performance evaluation of proposed design is illustrated in section IV. Section V concludes the paper.

2. Review Status Of Disaster Management

The review of literature was done in three different areas.

- The review on Pre-processing the data and representing them

- The review on predicting the impact of disaster using satellite map
- The review on communicating the victims during disaster within a short period

Pre-processing the data and representing them is explained by Igor Kononenko et al., [1].

Datta et al., have used feature selection algorithm for pre-processing. Here different features are selected and they are used to improve the accuracy of the classifier. They have used the classifier such as Naïve Bayes, HNB and NB Tree. [2]

Wei Zhao et al., claims that change detection method could be used for landslide mapping and they have tested with post-earth quake images. They have proposed a pixel based change detection method using the satellite image from Landsat-8 and have tested using 2015 Nepal Earth quake images[3]. Siti Nor Khuzaimah et al., have used neural network with three convolutional layers for disaster detection. They have tested by creating disaster detection training data patches and have shown an accuracy of 80 – 90 % [4]

Kamran Ali et al., proposed a disaster management communication networks. They have proposed a network architecture design. They have used the existing network infrastructure and have integrated with layer-based techniques. They have used cloud processing concepts. [5].

Flor Alvarez et al., have enumerated the risks and challenges that may arise during crisis and have provided practical guidelines for mitigating them during crises. [6]

Apart from these the social media can also be used for impact prediction as proposed by Soumya et. al [7]. They have made a new event detection approach by combining the linguistic feature of twitter. The traffic and its effect during disaster is proposed by Lin Qui et. al [8]. Here a parallel shortest path algorithm in road network is given

3. Impact Prediction Using KDD and SVM

The input data are details about the population, houses, hospitals, senior citizens, infants and Geographical map. The data is pre-processed using KDD algorithm [9],[10],[11]. It is shown in Fig 1.

In the Understanding phase, the data is analysed and represented in a tabular form. In the preparation phase, sampling and data cleaning operations are performed. This includes checking the completeness of data

records, correcting or removing for noise and missing values, etc. In the next phase the knowledge is discovered from the cleaned data. This knowledge includes the % of people in a particular area at any period of time which would be affected by the employment, climate vacation period etc. For eg., It is expected that the impact of disaster on life of people will be more during night time than day time.

The first level output is a map when zoomed will give the details of the selected area. A sample is shown in Fig 2

In the Understanding phase, the data is analyzed and represented in a tabular form. In the preparation phase, sampling and data cleaning operations are performed. This includes checking the completeness of data records, correcting or removing for noise and missing values, etc. In the next phase the knowledge is discovered from the cleaned data. This knowledge includes the % of people in a particular area at any period of time which would be affected by the employment, climate vacation period etc. For eg., It is expected that the impact of disaster on life of people will be more during night time than day time.

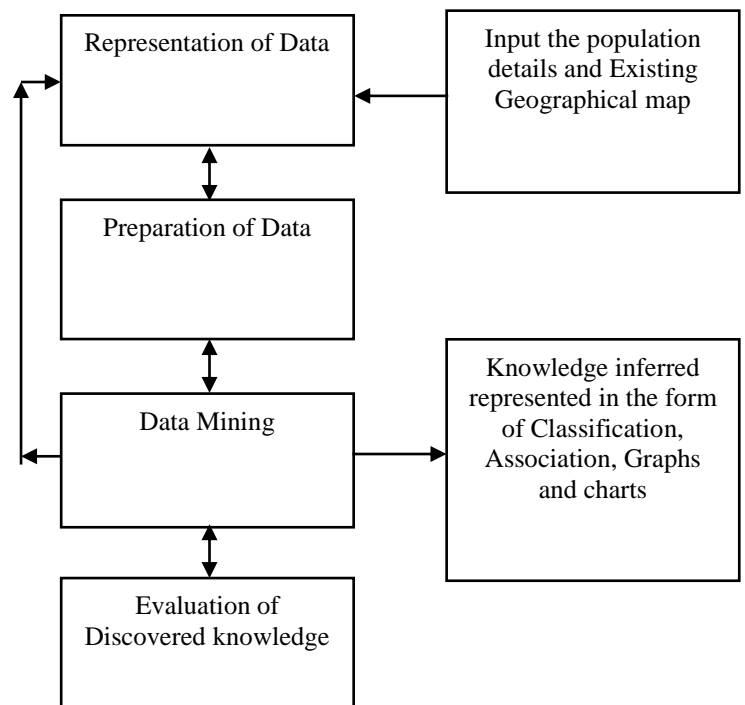


Fig 1: Data Pre-processing

The first level output is a map when zoomed will give the details of the selected area. A sample is shown in Fig 2



Fig 2: Sample details

Support vector machine(SVM) algorithm is applied in order to group the features as different categories. A hyperplane is formulated with a normal vector w and bias b . This separates two classes with required margin.

Let Samples be $X=\{(x-1,y-1),(x-2,y-2),\dots(x-n,y-n)\}$ where x_i is element of R^M , y_i lies between -1 and $+1$. The hyperplane is $w^T x+b=0$ with normal vector w and bias b in order to separate two margins.

The margin is calculated using equation (1)

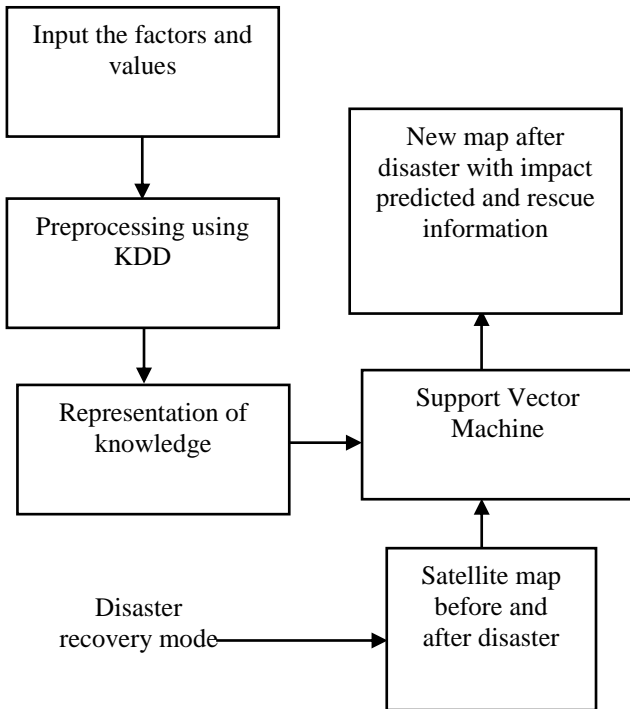


Fig. 3: Disaster prediction and management

$$P(a) = \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i,j=1}^N a_i a_j y_i y_j x_i^T x_j$$

(1)

such that $\sum_{i=1}^N a_i y_i = 0; 0 \leq a_i \leq C; C$

can be set based on the square kilometre of graph compared.

The output of this will have two information.

1. The impact of the disaster
2. The rescue path for the people

The disaster impact predicted could be used by the Government officials and NGOs for helping the needy

The rescue path can be used by the people in any area to save themselves.

The overall block diagram of disaster prediction and management is shown in Fig 3.

Fig 4: Pre and Post tsunami pictures

4 Evaluation

The evaluation of the work is done in three phases.

- The accuracy of the knowledge generated by KDD
- The accuracy of SVM in disaster impact prediction
- The accuracy of the algorithm in finding evacuation path

The knowledge represented by KDD gave accuracy of 95% since the data input is more clear.

The verification is done with Earth quake and Flood data.

Experimental setup for this work is done as follows

- Earth quake - 20 train data sets (200 patches per set) - 10 test data sets (200 patches per set)
- Flood experiment (Tsunami)- 5 train data sets (100 patches per set) - 3 test data sets (100 patches per set)
- For each set of data, we have minimum 2 aerial images which includes the input and output images: Sample input is shown in Figure 4.[12]
- The impact predicted was compared with the exact impact given by the government. It gave around 80 – 90 % accuracy.

The rescue path was tested with test data of earth quake and Flood images. The earth quake test data gave accuracy of 98 % and the flood data gave accuracy of 80%

5 Conclusion and Future Work

The result could be tested with few real examples only. The data set used were 10% real data and 90% generated data. Experimental evaluation gave a result around 90% overall.



The final product could be used as an application that can be downloaded by any user. Especially by the residents of disaster prone area.

When there is no power and no internet connection the work may not be useful for the residents. This could be overcome by placing wireless FM transmitter transmitting rescue information in a balloon and made fly over affected area.

The algorithm will not be at its best during night time and flood times since the map may not be clear. This can be taken as future work by receiving the immediate inputs from the residents using mobile or any other device and hence predicting the disaster impact as well as evacuation path.

References

- [1] Igor Kononenko, Matjaz Kukar, “*Machine Learning and Data Mining*“, 1st Edition, eBook ISBN: 9780857099440, Woodhead Publishing.
- [2] Datta H. Deshmukh, Tushar Ghorpade and Puja Padiya,(2015) “Improving classification using preprocessing and machine learning algorithms on NSL-KDD dataset”, *International Conference on Communication, Information & Computing Technology (ICCICT)*
- [3] Wei Zhao; Ainong Li; Xi Nan; Zhengjian Zhang; Guangbin Lei,(2017) “A pixel-based Post earthquake Landslides Mapping From Landsat-8 Data for the 2015 Nepal Earthquake Using a Pixel-Based Change Detection Method”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. Year: 2017,
- [4] Siti Nor Khuzaimah, Binti Amit, Soma Shiraishi, Tetsuo Inoshita and Yoshimitsu Aoki,(2016) “ Analysis of satellite images for disaster detection”, *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*
- [5] Kamran Ali, Huan X. Nguyen and Quoc-Tuan Vien, (2015) “Disaster management communication networks: Challenges and architecture design” , *Pervasive Computing and Communication Workshops (PerCom Workshops)*, *IEEE International Conference*
- [6] Flor Alvarez ; Matthias Hollick ; Paul Gardner-Stephen,(2016) “ Maintaining both availability and integrity of communications: Challenges and guidelines for data security and privacy during disasters and crises”, *Global Humanitarian Technology Conference (GHTC)*, October 2016
- [7] Soumaya Cherichi and Rim Faiz Larodec, “Using big data values to enhance social event detection pattern”, *13th International Conference of Computer Systems and Applications (AICCSA)*, 2016 IEEE/ACS 2016
- [8] Lin Qi and Markus Schneider, “ Trafforithm: A traffic-aware shortest path algorithm in real road networks with traffic influence factors”, (2015) *1st International Conference on Geographical Information Systems Theory, Applications and Management (GISTAM)*, 2015
- [9] Jayaram Hariharakrishnan, S. Mohanavalli, Srividya and K. B. Sundhara Kumar(2014), “Survey of pre-processing techniques for mining big data”, *International Conference on Computer, Communication and Signal Processing (ICCCSP)*, 2014
- [10] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, (1996), “The KDD Process for Extracting Useful Knowledge from Volumes of Data”, *COMMUNICATIONS OF THE ACM* November 1996/Vol. 39, No. 11
- [11] Yasuhide Okuyama,” Economic Modeling for Disaster Impact Analysis: Past, Present, and Future”,(2007), a Graduate School of International Relations, *International University of Japan , Niigata, Economic Systems Research* Vol. 19, No. 2, 115 –124, June 2007
- [12] Corneliu Octavian Dumitru, Shiyong Cui, Daniela Faur, and Mihai Datc,(2014), “Data Analytics for Rapid Mapping: Case Study of a Flooding Event in Germany and the Tsunami in Japan Using Very High Resolution SAR Images”, *IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing*, May 2014