**Building Damage Estimation**

**By**

**Muhammad Ali Zia**

**15K-2123**

**Saad Kamran**

**15K-2206**

**Sarim Balkhi**

**15K-2828**

**Supervisor**

**Mr. Muhammad Rafi**

**Assistant Professor**

**Computer Science Department**

**FAST – National University of Computer & Emerging**

**Sciences, Karachi**



**November, 2018**

**Version 1.0**

1. **Abstract**

Social media networks like Twitter or Facebook generate a lot of situational data during disastrous situations like flood or earthquake. It is already proven that the social media is quite fast on sharing the situational information/awareness. The rich content of shared media offers a lot of information. The damage can easily be seen through the pictures of the building shared after the disasters. Damage assessment is a vital task for many relief and welfare organization. Traditionally, damage assessment is slow and take a lot of time. The process is also very details on computing the actual loss of worth through many integrated cost components. The assessment of infrastructure damage is quite complicated. This Final Year Project (FYP) proposed a supervised approach to damaged assessment by utilizing computer vision techniques as well as semantically analyzing the textual representation of said and building a hybrid model of the two in order to provide the best estimate.

Much work has been done to assess the viability of using user generated data to help aid in this disastrous situations [14] [13] [12] [11] [8] [7] [6] [5]. Some focus on utilizing textual data generated by users [15] [13] where as others focus on using images posted by users and sometimes a combination of both [11] [9] [14]. This review focuses on getting a better understanding of these approaches.

1. **Introduction**

The emergence of social media sites specially twitter has been considered a useful platform for communication and an effective information delivery mechanism. During natural disasters or different violent activities, it provides the most recent information ranging from number of injuries to deaths, infrastructure damage and urgent rescue to people. All this information is in the form of text and images. Since to rescue the injured people by responding immediately social media sites plays a significant role to provide information in seconds or minutes.

In order to process useful information from these microblogging sites and help the government bodies to access and allocate resources accordingly we have developed a Natural Disaster analytics system for use by both governmental and non-governmental organizations. Mauricio and Libardo (2001)[16] developed a system (before the prevalence of social media platforms) which used fuzzy logic, systems engineering, and a rudimentary neural network architecture which was able to estimate damage caused by natural disasters to a high degree of accuracy but due to lack of crowdsourcing platforms there work had to be rely on human input and there work was quickly adapted for educational purposes, although their work was not sustainable it does show the potential of such systems, and this potential in combination with social media platforms has been explored by many ever since.

Many models that have been proposed classify the raw data into two classes i.e. relevant and irrelevant but our model has the capability to classify the data into three classes i.e. high, low and irrelevant. To deal with text and as well as images we seek to develop a hybrid model which could potentially classify based on both images and text in the same model. Due to these characteristics the model will have the ability to predict the damages caused on buildings by natural disasters. By the help of the predictions calculated by the model it can be decided that a building needs to be reconstructed as affected in natural disasters.

The project is divided in different stages firstly all the datasets has been collected for classification. All the images are processed through transfer learning and the literature is classified through CNN. After that the model will be trained and then it will predict the results for testing.

1. **Background:**

Natural disasters are an unpredictably predictable phenomenon which cannot be prevented and as such developing early warning systems is not always feasible using current technologies [1]. Their effects are devastating both in terms of human lives and to a countries economy and by extension to individuals living in it [2], Thus the importance of an adequate emergency management is acknowledged by most countries in the world. Disasters have always happened and they will continue to, what changes though, is global awareness of these disasters as well as the number of people affected. The terror attacks in New York, Spain and London as well as the Katrina hurricane, to name a few, are some recent examples of large scale emergencies. The cause of such emergencies could be natural disasters like flooding, earthquakes, volcanoes etc. Man-made disasters like terrorist attacks, industrial disasters, radiation contamination, etc. have increased dramatically in recent times.

Computer systems could facilitate all phases of Disaster Management. The aim is not to describe the functionality of every Disaster Management system available, but to investigate the way Human-Computer Interaction takes place in such systems. Especially the application and potential benefits of Neural Networks are considered.

Numerous tragic natural disasters threaten vulnerable areas of the world each year. Although there are some systems available to deal with the damages occur in natural disasters but the proposed system is more efficient than the previous systems as it uses semi-supervised deep learning approach and develop a hybrid model for classification of text and image as well.

1. **Role of messaging apps in these situations**

Much has been written concerning the value of using micro-blogging platforms data from crowds of non-professional participants during disasters. Data produced through micro-blogging platforms, e.g., Twitter, is seen as ubiquitous, rapid and accessible [3], and it is believed that empowering average citizens to become more situationally aware during disasters will allow them to coordinate and help themselves [4].

Time-critical analysis of social media data streams is important for many application areas. During the onset of a crisis situation, people use social media platforms to post situational updates, look for useful information, and ask for help [5]. X. Guan and C. Chen (2014) [17] attempted to quantify the evolution of disasters and thus demonstrate temporal–spatial patterns of Twitter activities as a means of establishing a pattern between twitter activity and damage from natural disaster.

This product mainly uses twitter as a source to collect raw data thus it has the capability to eliminate all the useless raw data. If a micro-blogging data consists of 6,000 images out of which only 2,000 images will be useful then the model based on neural network will classify all the images which will result in elimination of all the useless images. A hybrid model is also developed which uses all the images and text from social media sites in order to classify images as well as text in the same model and predict the damages caused on buildings by natural disasters. By the help of the predictions calculated by the model it can be decided that a building needs to be reconstructed as affected in natural disasters.

1. **Prevention:**

Rapid analysis of messages posted on microblogging platforms such as Twitter can help humanitarian organizations like the United Nations gain situational awareness, learn about urgent needs of affected people, critical infrastructure damage, medical emergencies, etc. at different locations, and decide on response actions accordingly [6].

In order to identify useful messages for humanitarian tasks one potential approach is to use supervised learning to automatically categorize each incoming message into one of the two classes i.e., relevant and irrelevant. In order to design the classification model, obtaining a large amount of labeled data is a challenging task, particularly during the first few hours of a crisis situation. However, access to abundant unlabeled data is possible under such time-critical situations, as hundreds of tweets arrive each minute. Moreover, one can rely on labeled data from past similar events. In such situations, semi supervised methods can provide effective ways to leverage unlabeled data in addition to labeled data [7].

The model uses semi-supervised deep learning approach to classify the incoming data into three classes i.e. high, low and irrelevant. In previous approaches mostly two classes have been implemented i.e. relevant and irrelevant but in this case the relevant class has been divided into two classes i.e. high damage and low damage. As the obtained labelled data always have a less amount as compared to unlabeled data that’s why semi-supervised method has been implemented to provide efficient and effective results.

1. **Monitor and assess the Damage:**

Studies have analyzed how big crisis data can be useful during major disasters so as to gain insight into the situation as it unfolds [10]. A number of systems have been developed to classify, extract, and summarize crisis relevant information from social media [3]. Cameron, et al., describe a platform for emergency situation awareness [12]. They classify interesting tweets using an SVM classifier. Verma, et al., use Naive Bayes and MaxEnt classifiers to find situational awareness tweets from several crises [13].

Thomas et al. [8] propose an automatic building damage assessment approach using pre-event and post-event aerial images. They consider an increase in the total number of edges appears on roof structures as a sign of damage. Yusuf et al. [9] evaluate the affected areas by calculating the difference between the brightness values in the pre and post-earthquake satellite images.

Greg et al. [11] automatically assess damage to buildings caused by natural disasters. They extract rich features from post-event images, as the only input source to their algorithm, and output a continuous value as a factor measuring damage severity rather than classifying damage into predefined categories.

This hybrid model uses Convolution Neural Network to classify the images and text into three different classes and predicts the damage caused on buildings during natural disasters using deep learning semi-supervised learning approach. It predicts the damage into two categories i.e. highly damaged and low damaged. By the help of predicted results, it can be decided whether to reconstruct or renovate the damaged buildings.

1. **Evaluation**

Throughout our review of present literature, we have seen a varying degree of methods employed by researches to evaluate the damage some chose to use pre and post image of a natural disaster whereas some choose to use post images [1] due to that fact that it is not possible to get pre event images for non-landmark buildings where as some opted to use satellite imagery to gather data [9]. Despite the variation in data one thing most of them had in common was the use of CNNs to classify images with some minor adjustments in their overall architecture

While the textual data had a varying degree of algorithms applied to them some researchers chose to use CNN whereas some coupled it with word embedding to remove linguistic barrier presented when analyzing data gathered throughout the world and this method seems to hold the most potential as it yielded the best result when compared to other methodology used in literatures we have seen

* 1. **Accuracy**

Although accuracies published in these papers should be treated as arbitrary figures when it comes to comparison as they have varying kind of dataset from satellite imagery to pre to post disaster datasets as these would not allow for a one to one comparison but regardless of that the accuracies do allow us to gain better understanding of best possible methodology which we should use. Xukun Li, Huaiyu Zhang, Doina Caragea and Muhammad Imran(2018) [14] had implemented a multipronged approach first they performed Damage Detection Map Evaluation using CNN which gave them a ~95.5% testing accuracy using a heatmap than they passed they results to another CNN to perform classification (Damage Assessment Value) which achieved a varying degrees of accuracy from ~53.3% to ~88.6%. Similarity Nia, Mori’s (2017). [1] model utilized CNN, their model contained three different pipelines and each pipeline is made up of one or two different ConvNets and they were able to achieve low error of about ~5% unfortunately they did not publish an overall accuracy. Based on the current research trend it is safe to assume that CNN provide the best result for work in this domain.

On the text front Caragea, Silvescu, Tapia (2016) used CNN to perform classification and achieved accuracies from ~77.61 to ~82.52 beating all other commonly used classification methods like SVM, Naïve approach, and ANN with ANN being the closest contender to CNN classifiers winning by a narrow margin of ~1-2% a similar trend can be observed in Nguyen, Mannai, Joty, Sajjad, Imran, Mitra’s (2016). [15] work in which they were able to achieve the highest level of accuracies through the use of CNN but their work goes one step beyond a rudimentary CNN they initialize the CNN with word embedding’s. The reason for departure from the more conventional methods of text classification by researchers in the domain is due to the fact that tweets are shorter compared to larger text bodies which are better suited for SVM and Naive Bayes Classifier, KNN.

1. **References**

[1] Karoon Rashedi Nia, Greg Mori (2017). Building Damage Assessment Using Deep Learning and Ground-Level Image Data

[2] Eduardo Cavallo, Sebastian Galiani, Ilan Noy, Juan Pantano (2013). Catastrophic Natural Disasters and Economic Growth

[3] Imran, M., Castillo, C., Diaz, F., Vieweg, S. (2013a). Processing Social Media Messages in Mass Emergency: A Survey. Journal of the ACM Computing Surveys (CSUR) Volume 47 Issue 4, July 2015, Article No. 67.

[4] Cornelia Caragea, Adrian Silvescu, Andrea H. Tapia (2016). Identifying Informative Messages in Disaster Events using Convolutional Neural Networks

[5] Dat Tien Nguyen, Kamela Ali Al Mannai, Shafiq Joty, Hassan Sajjad, Muhammad Imran, Prasenjit Mitra (2016). Rapid Classification of Crisis-Related Data on Social Networks using Convolutional Neural Networks

[6] S. Vieweg, C. Castillo, and M. Imran, “Integrating social media communications into the rapid assessment of sudden onset disasters,” in Social Informatics. Springer, 2014, pp. 444–461.

[7] Imran, Alam and Joty, “Graph Based Semi-Supervised Learning with Convolutional Neural Networks to Classify Crisis Related Tweets,” in ICWSM 2018

[8] J. T. A. K. K. Bowyer, “Towards a robust, automated hurricane damage assessment from high-resolution images,” in International Conference on Web Engineering (ICWE), 2011.

[9] Y. Yusuf, M. Matsuoka, and F. Yamazaki, “Damage assessment after 2001 gujarat earthquake using landsat-7 satellite images,” Journal of the Indian Society of Remote Sensing, vol. 29, no. 1-2, pp. 17–22, 2001.

[10] A. Acar and Y. Muraki, “Twitter for crisis communication: lessons learned from japan’s tsunami disaster,” International Journal of Web Based Communities, vol. 7, no. 3, pp. 392– 402, 2011.

[11] Karoon Rahedi Nia and Greg Mori, "Building Damage Assessment Using Deep Learning and Ground-Level Image Data," 2017 14th Conference on Computer and Robot Vision (CRV)

[12] M. A. Cameron, R. Power, B. Robinson, and J. Yin, “Emergency situation awareness from twitter for crisis management,” in Proceedings of the 21st international conference companion on World Wide Web. ACM, 2012, pp. 695–698.

[13] S. Verma, S. Vieweg, W. J. Corvey, L. Palen, J. H. Martin, M. Palmer, A. Schram, and K. M. Anderson, “Natural language processing to the rescue? extracting” situational awareness” tweets during mass emergency.” in ICWSM. Citeseer, 2011.

[14] Xukun Li, Huaiyu Zhang, Doina Caragea, Muhammad Imran (2018). Localizing and Quantifying Damage in Social Media Images

[15] Muhammad Imran, Prasenjit Mitra, and Carlos Castillo (2016) Twitter as a Lifeline: Human-annotated Twitter Corpora for NLP of Crisis-related Messages.

[16] Mauricio Sanchez-Silva, Libardo Gracia (2001). Earthquake Damage Assessment Based on Fuzzy Logic and Neural Networks

[17] Xiangyang Guan, Cynthia Chen (2014). Using social media data to understand and assess

disasters