# The Study of Quantitative Forecasting Model on City Emergency Incidents

Nan Gao, Xueming Shu, Jiting Xu, Biao Wen, Peng Chen, and Peng Wu

***Abstract*—Emergency incidents forecasting is quite significant to emergency response in Mega cities. In this paper, emergency incidents data were collected and data processing and analysis work were conducted. It is obvious that some rules exist when data was counted in different time spans (year and month). With discovered rules regression models were constructed and one of the current popular data mining software, Weka, was used to train and test the models. The results demonstrated that constructed linear regressed year-model and month-models fit the original data well (the MARE, mean average relative absolute error is less than 5%). The year and months' emergency incidents trend can be predicted based on this model. At the end of this article, some factors that produce model deviations were discussed from social activities perspective.**

***Index Terms*—Cross validation, emergency incidents, forecasting, linear regression.**

## I. INTRODUCTION

With the development of society and economy, modern society is facing more and more natural and disasters and social emergency incidents [1]. Particularly, along with the fast development of economy in China, it is quite significant to improve the systems and mechanisms for responding disasters and incidents to improve the mega city's comprehensive ability to prevent and handle emergencies. For example, in 2011, 44016 emergency incidents occurred in a southern China city, average 120 a day. From temporal perspective, the incidents seem to occur randomly over time. However, some sever incidents have its own regularity by time and it is necessary for people to discover the regularity and to make forecasting. As a matter of fact, in Mega city to response the emergency incidents is quite a challenge work because of both the human and property resources being in short. Thus, it is very important to construct a foresting-response scheme to predict the trend of emergency incidents and dynamically optimized and allocated the limited resources [2].

In this paper, some real data of emergency incidents in a mega city in south China were collected. The data was processed and analyzed and then forecasting models were constructed.

Manuscript received May 9, 2013; revised July 5, 2013. This paper was sponsored by National Science and Technology Support Program (2011BAK07B01) and Multi-dimension Data analysis and Decision-Support to Mega City's Emergency Response System(2011BAK07B04).

Nan Gao and Jiting Xu are with Computer Science and Engineering Department, University of South Carolina, Columbia, SC 29205 USA (e-mail:gaon@email.sc.edu, xu57@email.sc.edu).

Xueming Shu is with Institute for Public Safety Research, Tsinghua University, Haidian District, Beijing, 100084 (e-mail: shuxm@tsinghua.edu.cn).

## II. METHODS

In Statistics, Linear regression is a method to predict the relationship between dependent variable and one or more explanatory variables denoted by $X$. In linear regression, data are modeled using linear predictor functions, and model parameters are estimated from the data. Most commonly, linear regression is a model where the mean of $Y$ is estimated based on the observed value of $X$ [3]. Like the other forms of regression analysis, what linear regression focuses is the conditional probability distribution of dependent variable given explanatory variable, rather than on the joint probability distribution of dependent and explanatory variable, which is the domain of multivariate analysis.

Given a dependent variable $Y$ and an explanatory variable $X$, a linear regression model assumes that the relationship between the dependent variable $Y$ and the explanatory variable X is linear, so there is a linear equation:

$$Y=\alpha+\beta X+\varepsilon$$

where $\alpha$ is constant, $\beta$ is the coefficient, $\varepsilon$ is residual.

Linear regression model is the first one that had been strictly researched and broadly used in practical (Draper *et al.*, 1998 [4]). Generally, linear regression model is frequently approximated or estimated with least squared method.

## III. WEKA

Weka (Homes, 1954; Witten *et al.*, 2011) [5], which is the short of Waikato Environment for Knowledge Analysis, is freely used software in machine mining and data mining. The software is coded in JAVA which enforces it being used in multi-platforms. As being a open data mining plat form, Weka integrates lots of algorithms like classification, regression, clustering, correlation analysis and visualization in new interactive interface.

In our work, Weka was used to build and test the model based on the data, which can be employed to explore the trend of the number of emergency accidents. To make the best use of the dataset, training datasets and test datasets were constructed by the advantage of 10-fold cross validation. Then the model were constructed, trained, and tested [6].

## IV. DATA ANALYSIS AND MODEL

The data was collected from a mega city's emergency center, it was provided in an Oracle database. The temporal attribute of the data starts at 2002 and ends at 2011. There is 97000 recordings in the database. There are 20 attributes, but only four of them were used in this paper (I am not sure about this)due to too much missing values for the rest attributes, they are date (year/month/day), time (hour/minute/second), incidents type (police/fire/first aid/traffic), location.

In the database, as the emergency incidents recordings are the most complete and typical attributes as well as it has high correlation with police, fire, and traffic recording, therefore the first aid recordings were employed to construct the model. Two steps were performed to accomplish this task. First, the yearly first aid recordings variation is studied and year forecasting model was constructed. Then the data in every month in past ten years was studied and then month-forecasting model was constructed. With the two models, forecasting by year and month could be achieved.

Table I displays the count of incidents from 2002 to 2011. It can be seen that the traffic and police data are not complete. Since the first aid incident recordings are good indicators to the response of the emergency management, thus the first aid data was used to construct the model to analyze and forecast the emergency incidents.

TABLE I: EMERGENCY RESPONSES TABLE OF 2002 TO 2012

| YEAR | EMERGENCY INCIDENTS | POLICE | FIRE | TRAFFIC | TOTAL |
|---|---|---|---|---|---|
| 2002 | 11629 | 52011 | 1388 | 14987 | 80015 |
| 2003 | 17518 | 80647 | 2066 | 20260 | 120497 |
| 2004 | 19592 | 4293 | 2432 | 287 | 26604 |
| 2005 | 21736 | 2172 | 2803 | | 26711 |
| 2006 | 24880 | 1219 | 3853 | | 29952 |
| 2007 | 29586 | 820 | 4066 | | 34472 |
| 2008 | 33673 | 1277 | 4687 | | 39637 |
| 2009 | 35734 | 719 | 4834 | | 41287 |
| 2010 | 40148 | 267 | 4391 | | 44806 |
| 2011 | 44106 | 325 | 5030 | | 49371 |

10-Cross-validation method was used to train and test the model. It randomly partitions the sample of data into 10 complementary subsets. Of the 10 subsets, a single subset was retained as the validation data for testing the model, and the remaining 9 subsets were used to train the model. And we average the results from the 10 folds to obtain the final estimation. Therefore, all observed data were used for both training and validation, and each observed data is used for validation exactly once. The model was tested by computing its squared error. This process was iterated until all sample data was used to make test. Finally, the squared error was summed to be PRESS (predicted error sum of squares).

### A. Forecast Model

#### 1) Year-forecasting model

First of all, the count of first aid data was accumulated into each year to create a ten-year series. In Weka, the data was cleaned and regressed with linear fitting and 10-fold cross validation method. The regressed model was shown below:

$$y = 3461.33 \times (\text{year} - 2001) + 8813.87$$

where $y$ is dependent variable, which means the forecasted first aid count. Year denotes the explanatory variable.

The corrected coefficient is 0.9939 via cross-validation. The relative error is 9%. Fig. 1 demonstrated the real data and the forecasting model. According to the figure, the model shows the variation trend of first aid by year, so it was satisfied in forecasting future first aid count.

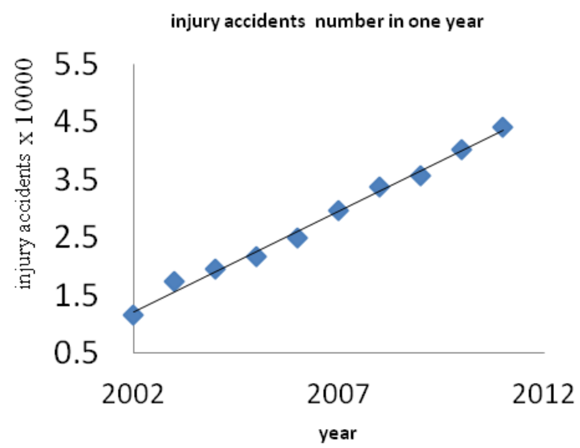


Fig. 1. Yearly emergency incidents real data and prediction data.

#### 2) Month forecast model

For the month, firstly, the 10-year first aid recording data was summed up into months and box plots method was used to eliminate the outliers. Also in Weka, linear trend fitting and 10-fold cross validation method was used, and eventually 12 linear regression models were created. The results are displayed in Table II.

After testing, it was discovered that the corrected coefficients to the model were more than 0.95, and relative errors were less than 13%. It could be seen that the models fitted the data well and can be used in trend forecasting.

TABLE II: MONTH-FORECASTING MODEL

| **Mon** | **Linear Prediction Model** | **Mon.** | **Linear Prediction Model** |
|---|---|---|---|
| Jan. | 327.6*(year-2001)+704.5 | July | 301.8*(year-2001)+ 698.8 |
| Feb. | 259.5*(year-2001)+ 41.0 | Aug. | 290.0*(year-2001)+ 683.4 |
| Mar. | 283.8*(year-2001)+ 603.4 | Sep. | 286.9*(year-2001)+ 751.8 |
| Apr. | 251.5*(year-2001)+ 786.4 | Oct. | 286.7*(year-2001)+ 864.6 |
| May | 237.2*(year-2001)+ 1022 | Nov. | 309.1*(year-2001)+ 691.9 |
| June | 282.3*(year-2001)+ 757.8 | Dec. | 290.5*(year-2001)+ 1002 |

### B. Model Analysis

The models were analyzed further combined with background events.

For the year-forecasting model, it could be discovered that the largest error appeared in 2003, when atypical pneumonia occurred and spread out in the whole country. Therefore it is indicated that important health events had impact on emergency responses to the mega city, especially to the first aid departments.

For the month-forecasting model, it was observed that there was a great deviation between real data and forecasted data on February and August in 2008. If dated back to that period it can be inferred that National People's Congress meeting and Olympic Games were held at that time. Thus it is indicated that some social events might have huge influence to the emergency response, so these events should be included into the model to make a better forecasting. Based on these phenomenons, it can be seen that the count of incidents fluctuate through the year and this feature correlates significantly with social activities. So, for emergency departments, human and property resources should be arranged and allocated properly based on these certain social factors (see Fig. 2).

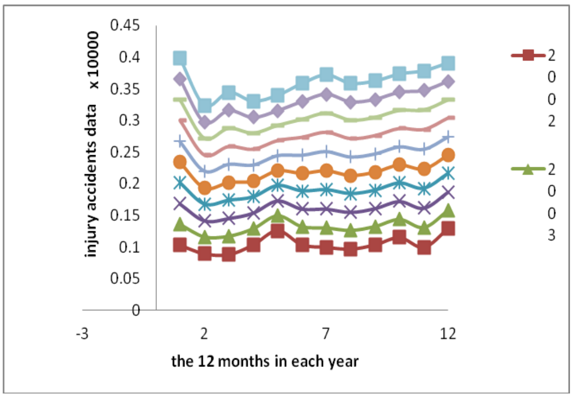


Fig. 2. The lines of the number of each month emergency incidents in 10 year.

## Acknowlegement

This paper was sponsored by National Science and Technology Support Program (2011BAK07B01, 2011BAK07B04) andMulti-dimension Data analysis and Decision-Support to Mega City's Emergency Response System (2011BAK07B04-B)

## References

[1] M. Felson and E. Poulsen, "Simple indicators of crime by time of day," *International Journal of Forecasting*, vol. 19, no. 4, pp. 595-601, 2003.

[2] L. Wang and B.-Q. Zhao, "Study to the interfere time, resource and time series of multi-response," *Social scientist*, vol. 2009, no. 6, pp.107-112.

[3] J. Cohen, P. Cohen, S. G.West, and L. S. Aiken, "Applied multiple regression/correlation analysis for the behavioral sciences," *Hillsdale*, NJ: Lawrence Erlbaum Associates, 2003.

[4] N. R. Draper and H. Smith, *Applied Regression Analysis*, Wiley Series in Probability and Statistics, 1998.

[5] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical machine learning tools and techniques*, 3rd Edition, San Francisco: Morgan Kaufmann, 2011.

[6] G. Holmes, A. Donkin, and I. H. Witten "Weka: A machine learning workbench," in *Proc. Second Australia and New Zealand Conference on Intelligent Information Systems, Brisbane*, Australia, 1994.

**Nan Gao** was born in Hefei, in 1983. She got her master degree in Tinghua University, China. Now she is a Ph.D candidate in University of South Carolina, Columbia. She focuses on data mining, data analysis and pattern recognition.

**Xueming Shu** was born in Sep. 1970, in Anhui, China. She is a Professor at Qinghua University, BJ, China.

**Jiting Xu** was born in Dec. 1986, Shandong, China. Now he is a M.S. candidate in University of South Carolina, Columbia. He focuses on data mining, data analysis and pattern recognition.

**Bian Wen** was born in Feb. 1967, in Guangxi, China. He is a senior engineer and Academic leader. He is interested in the application of information technology in emergency management.

**Peng Chen** was born in Feb. 1981, China. He is a Ph.D in safety science, lecturer in people's public surety university of China, emergency response.

**Peng Wu** was born in Feb. 1975, in Jianxi, China. He received his MBA. He is interested in public safety.