A GIS-based Demand Forecast using Machine Learning for Emergency Medical Services

Albert Y. Chen[1,] and Tsung-Yu Lu[2]

[1]Assistant Professor, Department of Civil Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617, Taiwan; PH +886-2-33664255; FAX +886-2-23639990; email: albertchen@ntu.edu.tw
[2]Graduate Research Assistant, Department of Civil Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617, Taiwan; PH +886-2-33664255; FAX +886-2-23639990; email: r01521529@ntu.edu.tw (*Corresponding Author)

ABSTRACT

The objective for pre-hospital Emergency Medical Service (EMS) is to reach to, pick up, and deliver patients efficiently. By increasing the operational efficiency, the survival rate of major trauma patients could potentially be improved. In this research, the authors applied Moving Average, Artificial Neural Network, Linear Regression, and Support Vector Machine for the forecast of pre-hospital emergency medical demand. The results from these approaches, as a reference, could be used for pre-allocation of ambulances. The training of the models as well as the validation is conducted with data collected from the New Taipei City EMS. In order to represent the performance of forecast, the authors introduced Geographic Information System (GIS) to manage and visualize the spatial distribution of forecast demand and error. Moreover, the authors built a flexible model, which organizes demand data with the user given size of area and size of time step. With the easy use of model and acceptable prediction performance, the research outcome has its potential to be applied to the current practice.

INTRODUCTION

With the advance of computer hardware and development of information technology in the past few decades, the concept of big data and data analytics have risen. As machine learning has shown good potential for the analytics of big data, the efficiency of Emergency Medical Service (EMS) operations has the potential to be increased using data driven approaches. The authors hope by taking this approach, information to make better decisions could be revealed and the survival rate of emergency patients could be increased.

The response time is in general considered as a performance index for EMS operations. The definition of response time is the time from an operator receives the EMS call until the ambulance arrives at the scene. The following statistics shows the importance of the response time, according to the Out of Hospital Cardiac Arrest (OHCA) cases from New Taipei City in 2010. The response time has an impact on

survival, and by shortening the response time there is chance to improve patients' survival rate.

Table 1. Survival Rate in each time bucket

Response Time	Number of Events	72 Hour Survival	Survival Rate
0 - 3 min	172	30	17.44%
3 - 6 min	1243	129	10.38%
6 - 10 min	788	91	11.55%
> 10 min	199	12	6.03%

This study aims to study how to shorten the response time in order to improve the efficiency and provide the pre-hospital EMS timely. The goal is to have accurate predictions such that EMS resources can be allocate more reasonably and more efficiently (Fitch, J. 2005). The authors applied Moving Average (MV), Artificial Neural Network (ANN), Linear Regression (LR), and Support Vector Machine (SVM) to forecast the location and number of EMS requests in the near time step. The demand prediction and its errors in each region are visualized in GIS.

CURRENT LITERATURE

In recent years, GIS has assisted the EMS research and practice to understand the distribution of the data, which enabled examination of the existing spatial allocation of medical resources (Lerner et al. 1999; Warden. 2007; Ong et al. 2008). Patuwo et al. (1998) summarized the advantages and disadvantages of the ANN and suggested the suitable situations, such as large data sets, problems with nonlinear structure, and the multivariate time series forecasting problems.

ANNs have been applied to various research areas. Its ability to self learn, to generalize, and to model nonlinear relationships, and multilayer feed forward neural network (MFNN) are known as a general way at building relationships between inputs and known outputs (Smith & Gupta, 2000). Setzler et al. (2009) applied ANN to predict EMS call volume. They suggested that ANN do not require assumption of the properties of data, and it can model complex data pattern which are difficult to observe or describe. In their paper, Setzler et al. (2009) showed that the EMS demand forecast should consider its spatial and temporal attributes. Further, where and when the EMS call will occur in practice is a major concern for EMS practitioners (Cheng et al. 1997).

There have been studies in the field of emergency medical service using Support Vector Machines (SVMs) (Nasiri et al. 2009, Chiu et al. 2010, Mehta et al. 2007). All of which adapted SVMs on classification problems, but few study discuss the potential of the SVM's regression capabilities. This study focuses on how to use SVMs regression to predict the EMS call volume in comparison to other models. Additionally, integrated with GIS, the forecast results can be visualized more intuitively.

OBJECTIVE

The objective of this research is to reduce the response time for ambulances. Machine Learning models are compared for demand forecast of pre-hospital EMS. The models facilitate emergency planning and decision-making by taking into consideration the temporal and spatial characteristics with the support of GIS for spatial analysis and visualization. The goal is to pre-allocate ambulances at locations with high probability of demand. This is expected to shorten the response time for patients' stabilization and treatment.

APPROACH

The proposed approach shown in Figure 1 is divided into two components. The first is GIS and the other Machine Learning (ML). The process of the approach starts in the GIS component with the data management block that produces a grid layer for the study area and calculates demand in each grid given the historical data. Considering the size of the studying area may change, the model is designed so that the grid size is given as input by the user. After setting the size of grid, the demand in each grid is calculated and the grids serve as input for the algorithm block. By breaking the study area into grids creates many smaller forecasting problems, and each grid has its own forecast model to capture the trend of EMS volume variation in each grid.

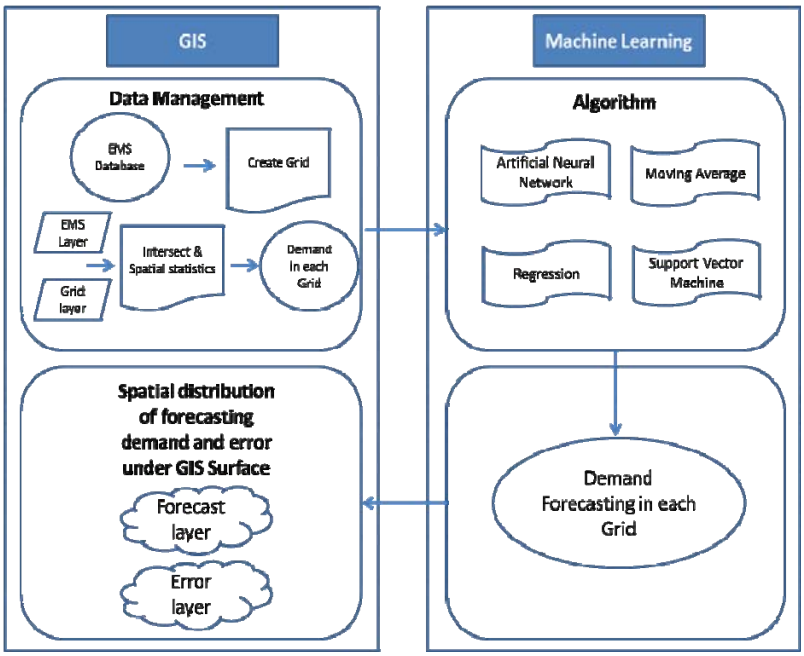

Figure 1. Overview of Data and Analyses in the EMS Assessment Process

Next, in the algorithms block are four algorithms including ANN, MA, LR and SVM for regression (SVR). Each grid is trained by using each algorithm and produces a forecasting result. Setzler et al. (2009) discuss how to apply the MFNN in EMS forecast, and the authors continue to follow the model of MFNN and attempt to improve model. In their research, they used four temporal factors without the attribute

of year. The authors believe this variable is necessary for the modeling so that the ANN Model can better capture the demand change over the years, and it may improve forecast accuracy. For SVR in this study, model selection was performed to optimize the parameters and all of four algorithms use the method of batch training.

To better visualize the EMS call volume, the forecast results from the algorithm block is send back to the GIS component. The authors believe the visualization provides better information to EMS managers with spatial distribution of demand in their jurisdiction.

CASE STUDY

Data from New Taipei City was collected from the years 2010 to 2012. The average EMS incidents are about 140,000 in each year. The authors make a daily forecasts using 34 months as training data and the last two months for validation. The Preliminary results are presented as the following. Figure 2(a) and Figure 2(b) shows the actual demand and the forecasted demand using SVR in a particular day. In Figure 2(b), SVR captures the feature both in the area of high EMS call volume and low volume. The way the demands are divided spatially are based on the grids shown in Figure 3, and the authors choose the 3km x 3km as grid size.

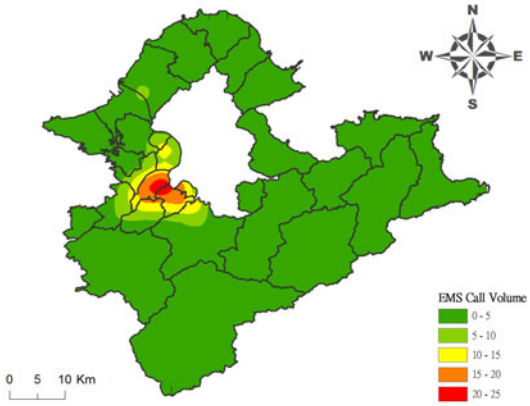

Figure 2(a). True Demand

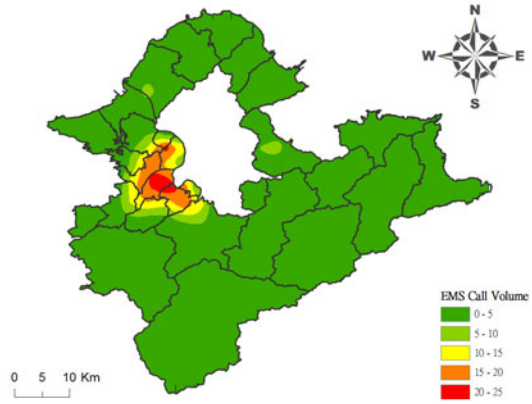

Figure 2(b). Predict Demand by SVM

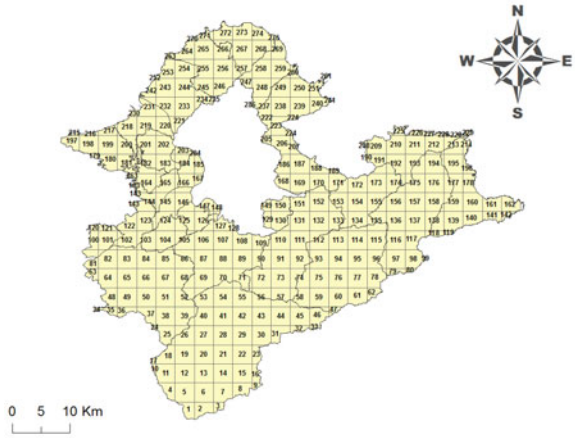

Figure 3. Grid ID of New Taipei City

Each model produces the daily forecast for the last two months. Once a forecast for a day is complete, the actual information is updated into the training set and the models are retrained using batch training. In other words, the training method continues to add data from the previous time step for the prediction of the next period.

The Mean Absolute Percentage Error (MAPE) is defined by the absolute value of difference between actual value and forecast value divided by the actual value, and then summed every time periods and divided again by number of time periods. The drawback of this index is if the actual value is zero, which causes a division by zero. In this study, there are many grids with no demand in the day. Once the model predicts a non-zero value, the MAPE will have an infinite value due to the division of zero. In such case, the MAPE is adjusted to 100% in this work. Based on Lewis' MAPE definition (Lewis, 1982), if the MAPE is less than 10%, the forecast is considered high accuracy, MAPE is more than 10% and less than 20%, it is considered good forecasting, the MAPE is greater than 20% and less than 50%, the forecast is reasonable and when MAPE is more than 50%, it is considered inaccurate forecasting. The MAPE is imported in GIS for visualization, as shown in Figure 4(a), Figure 4(b), Figure 5(a) and Figure 5(b).

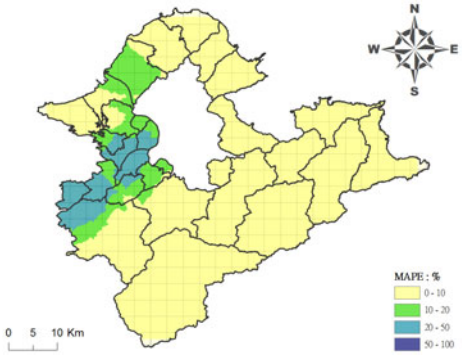

Figure 4(a). MAPE of ANN

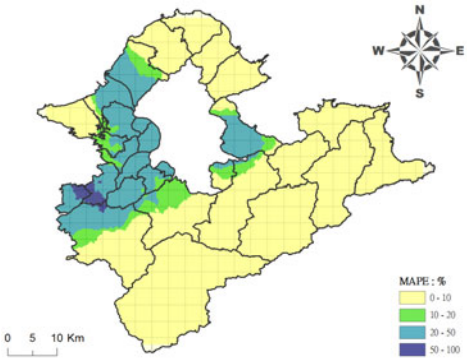

Figure 4(b). MAPE of MA

DISSCUSSION

The MAPE maps reveal the forecasting accuracy for each grid. From the four figures (Figures 4(a), 4(b), 5(a) and 5(b)), it seems that the ANN has better accuracy than other models.

Figure 3 reveal the grid layer of New Taipei City, and the grid id are shown. The authors analyzed the EMS data collected from the years 2010 to 2012, and found some grids have great EMS demand every year than other grid, those the grids with high EMS call volume as just described. In Table 2, the authors choose the grids with high EMS call volume for further discussion. SVR has better performance in most of these grids with relatively higher demand. Although in Figure 7 there appears to have many grids with MAPE more than 50% for SVR, it is caused by the division by zero of MAPE when there is zero actual demand as mentioned previously.

The proposed approach for pre-hospital emergency healthcare is an ongoing research. The model selection used in SVR cause better performance than other

models (see Table 2). In the future, each model will undergo a model selection step and a most suitable model will be selected for each grid.

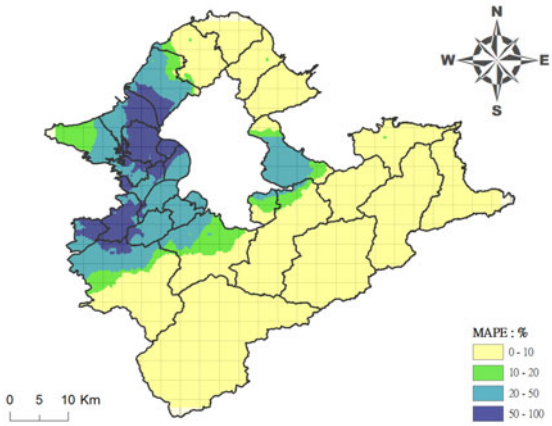

Figure 5(a). MAPE of LR

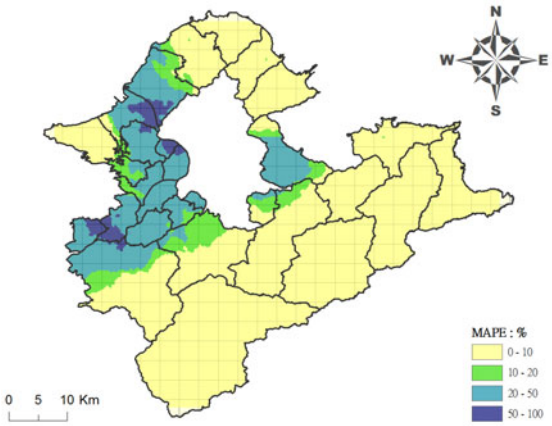

Figure 5(b). MAPE of SVR

Although the SVR with model selection in general have better performance than other models as seen in Table 2, some models outperform others in particular grids. For example in grid No.125 and No.143, the best model is MA. Each grid should have its own model that best represents it. If each grid has its own best model, it is believed that the overall system performance will be improved. After the model selection for all models for all grids, the best suitable model for the grid is determined.

Table 2. MAPE (%) of Four Algorithms

Grid ID	ANN	Regression	MA	SVR
No.122	36.37	35.14	36.84	31.63
No.123	35.08	33.42	31.52	26.22
No.124	46.45	35.27	33.76	32.08
No.125	42.86	28.34	21.55	29.20
No.143	70.30	55.80	30.04	39.15
No.144	30.74	27.81	21.97	21.60
No.145	27.49	19.86	19.22	16.15

Moreover, the training method in this study is batch training. There is the potential problem of using all data from training set for each iteration. In the case when the training set is large, training of models will consume a great amount of time. Online training has its potential to be adapted for this approach. It produces close to optimal solution relatively fast. The slight error is acceptable in practice, with the gain in execution time reduction. As the study emphasizes on real-time or efficient forecast of EMS demand, online training is a more appropriate approach.

CONCLUSION

The paper presents the demand forecast of four models for the pre-hospital EMS. Encouraging results are shown based on the case study presented. Models that give good forecasting results have been presented. The methodology proposed in this work has its potential to be applied given the historical data of EMS services. The proposed approach for pre-hospital emergency healthcare is an ongoing research, and future work is presented in the Discussion section. Although different cities have different characteristics of geographical layout, traffic conditions, and urban usage, as long as there is the historical data, this approach has its potential to be deployed. According to the demand forecasting, EMS managers can allocate the ambulance to the area with the higher probability of accidents in advanced. When the incident occurs, it is expected that ambulances will have a shorter response time for treatment.

ACKNOWLEDGEMENTS

The authors thank the Institute of Transportation, Minister of Transportation, Taiwan for the spatial data provided, and the EMS data from the Fire Department of the New Taipei City Government. Additionally, the authors thank the National Research Council of Taiwan for grants NSC 102-2627-M-002-016 and NSC 102-2221-E-002-119 that have supported this work.

REFERENCES

Chiu, C. C., Yeh, S. J., and Li. T. Y. (2010). "Classification of Diabetics with Various Degrees of Autonomic Neuropathy Based on Linear and Nonlinear Features Using Support Vector Machine." *Second International Conference, ICMB 2010*, 42-51.

Chang C. C, Lin CJ. LIBSVM: a library for support vector machines [Internet]. [cited at 2010 Oct 29]. Available from: http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

Fitch, J. (2005). "Response times: myths, measurement and management." *Journal of Emergency Medical Services*, 30(9), 47–56.

Ko, C. I. (2008). "Spatial Distribution and Influence Factors of Cross-District Transports among major trauma in emergency medical services system." Master Thesis, National Taiwan University.

Lewis, E.B. (1982). "Control of body segment differentiation in Drosophila by the bithorax gene complex," *Embryonic Development, Part A: Genetics Aspects*, Edited by Burger, M.M. and R. Weber. Alan R. Liss, New York, 269-288.

Lerner, E. B., Billitier, A. J. 4th., Sikora, J., & Moscati, R. M. (1999). "Use of a geographic information system to determine appropriate means of trauma patient transport," *Academic Emergency Medicine, vol*. 6, pp. 1127–1133.

Liu, H.-H. and Chen, A. Y. (2012). "Emergency Medical Dispatch: A Case Study of New Taipei City." Presented at 12th International Conference on Construction Application of Virtual Reality (CONVR 2012). Taipei, Taiwan.

Liu , H.-H., Chen, A. Y., Dai, C.-Y., and Sun, W.-Z. (2013). "Physical Infrastructure Assessment for Emergency Medical Response." (submitted), *ASCE, Journal of Computing in Civil Engineering*, Reston, VA, USA.

Mehta, S. S., Lingayat, N. S. (2007). "Support Vector Machine for Cardiac Beat Detection in Single Lead Electrocardiogram. " *IAENG International Journal of Applied Mathematics,* 36:2, IJAM_36_2_4.

Nasiri, J. A., Naghibzadeh, M., Yazdi, H.S., Naghibzadeh, B. (2009). "ECG Arrhythmia Classification with Support Vector Machines and Genetic Algorithm." *Computer Modeling and Simulation, Third UKSim European Symposium,* 187 – 192.

Ong, M. E. H., Tan, E. H., Yan, X. Anushia, p., Lim, S. H., Leong, B. S., Ong, V. Y. K., Tiah, L., Yap, S., Overtib, J., & Anantharaman, V. (2008). "An observational study describing the geographic-time distribution of cardiac arrests in Singapore: What is the utility of geographic information systems for planning public access defibrillation? (PADS Phase I)," *Resuscitation*, vol. 76, pp. 388–396.

Pathak, E. B., Reader, S., Tanner, J. P., & Casper, M. L. (2011). "Spatial clustering of non-transported cardiac decedents: the results of a point pattern analysis and an inquiry into social environmental correlates," *International Journal of Health Geographics*, 10:46.

Setzler, H., Saydam, C., & Park, S. (2009). "EMS call volume predictions: A comparative study. " *Comput. Oper. Res*. 36 1843–1851.

Smith, K. A., & Gupta, J. N. D. (2000). "Neural networks in business: techniques and applications for the operations researcher." *Computers and Operations Research*, 27(11), 1023–1044.

Warden, C. R., Daya, M., & LeGrady, L. A. (2007). "Using geographic information systems to evaluate cardiac arrest survival," *Prehospital Emergency Care, vol*. 11, pp. 19–24.

Zaki, A. S., Cheng, H. K., & Parker, B. R. (1997). "A simulation model for the analysis and management of an emergency service system." *Socio-Economic Planning Sciences*, 31(3), 173-189.

Zhang, G., Patuwo, E. B., & Hu, M. Y. (1998). "Forecasting with artificial neural networks: the state of the art." *International Journal of Forecasting*, 14, 35–62.