University of Tabriz

Journal of Civil and Environmental Engineering
Volume 48 (2018), Issue 1 (Spring), 67-75

EXTENDED ABSTRACTS

Performance Assessment of Computational Intelligence Techniques in Solid Waste Generation Forecasting (Case Study)

Mohammad Minousepehr, Mohammad Reza Alizadeh, Nasser Talebbeydokhti*

Department of Civil and Environmental Engineering, Shiraz University, Shiraz, Iran

Received: 04 August 2016; **Accepted:** 289 January 2017

Keywords:
Solid waste Generation forecasting, Multi-layer perceptron, Support vector machines, M5P model trees.

1. Introduction

Knowing the quantity of generated solid waste play a very significant role in solid waste management programs in a region. Due to lack of measured data as well as unavoidable errors in measurements, assessment of volume of generated solid waste is always challenging. Also, field measurement and continues monitoring of the volume of solid waste is usually costly, difficult and time-consuming. Accurate prediction of solid waste generation can be regarded as a key factor in future solid waste management system planning. Conventional forecasting methods in solid waste generation forecasting frequently use the demographic and socioeconomic factors in a per capita basis. In most cases, insufficient funds, the limited measuring equipment, lack of appropriate management systems and due to the lack of recorded data for the volume of generated solid waste cause many problems in integrated solid waste systems management (Dyson and Chang, 2005). In this study, three computational intelligence techniques including M5P model trees, support vector machines (SVM) and multi-layer perceptron (MLP) artificial neural network are utilized to predict solid waste generation in Hormozgan Province, Iran. After a sensitivity analysis, four more influential factors including elevation, population, urban development index (measures the level of development in cities based on infrastructure, the municipality established year, the metropolitan area, population, city product and income, health and education) and the frequency of garbage collection were considered in developing models. The performance of proposed models in solid waste generation forecasting are assessed via different error evaluation indices and finally the results are compared.

2. Methodology

2.1. Case study and data

The study area includes about 220 cities and Rural areas with an approximate length of 900 km in Hormozgan, a coastal province in south of Iran. This area begins with Koshkonar in the most western point of the area that is adjacent to the Persian Gulf and ends to the port of Jask, the easternmost point of the province which is adjacent to Oman Sea. In this study, due to the lack of recorded data for the volume of generated solid waste in the region, the per capita municipal solid waste generation was determined in different parts of the study area. By direct sampling of the generated waste in the major sources of waste production including residential and commercial areas in rural and urban areas during the four seasons of the years 2013-2014 (March 2013 to February 2014), a dataset for per capita solid waste generation was obtained.

* Corresponding Author
E-mail addresses: m_minousepehr@yahoo.com (Mohammad Minousepehr), alizadeh.mohamadreza@yahoo.com (Mohammad Reza Alizadeh), nassertaleb@gmail.com (Nasser Talebbeydokhti).

2.2. MLP

The neural models are basically based on the perceived process of the human brain. The artificial model of the brain is known as ANN. The theory and mathematical of Neural networks are mathematical models, which can assess the relation among inputs and outputs of a physical system (even if it is complex and nonlinear) through a network of attached nodes. ANN structure is called architecture, which is in a form arranged by neurons that are called layers. ANN customary architecture is composed of three layers: input layer, hidden layer and output layer. In our research, one hidden layer feedforward ANN was used for the SWG prediction. The Levenberg–Marquardt (L–M) algorithm was chosen for training the networks. In addition, feed-forward ANN was trained by the standard back propagation.

2.3. SVM

Kernel-based techniques such as support vector machines represent a major development in machine learning algorithms. Support vector machines (SVM) are a group of supervised learning methods that can be applied for classification or regression. Support vector machines represent an extension to nonlinear models of the generalized portrait algorithm developed by Vladimir Vapnik. The SVM algorithm is based on the statistical learning theory (Vapnik, 1995).

2.4. Model trees

M5P model trees were first introduced by Quinlan (1992) and then the idea was reconstructed and improved in a system called M5P by Wang and Witten (1997). An M5P model tree is an effective learning method for predicting real values. Model trees, like regression trees, are efficient for large data sets. However, model trees are generally much smaller and are more accurate than regression trees (Quinlan, 1992). The M5 model trees algorithm builds a regression tree by recessive splitting of the sample space to minimize the intra-subset variability in the values down from the root via branch to the node. The splitting processes continue up to the point that the output values fluctuate only slightly, or only a few samples remain. Remaining samples are output values of all samples that have reached the nodes. The variability is determined by standard deviation of the values that reach that node from the root through the branch. The standard deviation reduction (SDR) is calculated as follows:

$$SDR = sd(T) - \sum_{i} \frac{|T_i|}{|T|} \times sd(T_i) \tag{1}$$

Where T is the set of examples that reach the node, T_i the sets that are resulted from splitting the node according to the chosen attribute and *sd* is the standard deviation.

2.5. Evaluation of model accuracy

To evaluate the performance of developed models in forecasting values of per capita waste generation, four indices are assessed: root mean square error (RMSE) or (MAPE), mean absolute error (MAE), correlation coefficient (CC) and mean absolute relative error (MARE). The accuracy of the model is determined according to these criterias.

$$RMSE = \sqrt{\frac{1}{n}\sum (x_i - y_i)^2} \tag{2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |(x_i - y_i)| \tag{3}$$

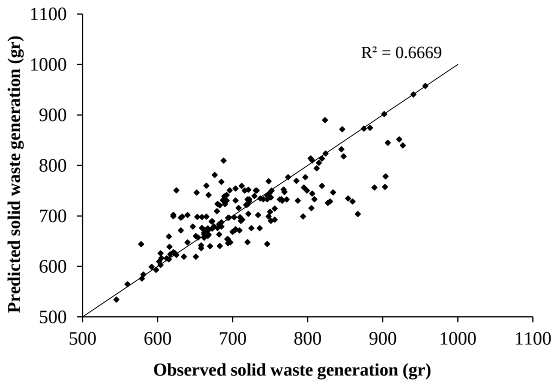
$$CC = \frac{\left(\frac{1}{n}\right)\left[(x_i - \mu_x)^T (y_i - \mu_y)\right]}{\sqrt{(1/n)(x_i - \mu_x)^2}\sqrt{(1/n)(y_i - \mu_y)^2}} \tag{4}$$

$$MARE = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{(x_i - y_i)}{x_i} \times 100 \right| \tag{5}$$

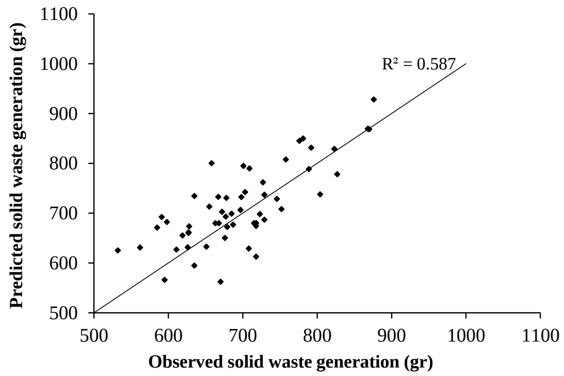
In all formulas, the x_i represents the observation, the y_i represents the predicted value, n the number of observations, μ_x the mean of x and μ_y the mean of y.

3. Results and discussion

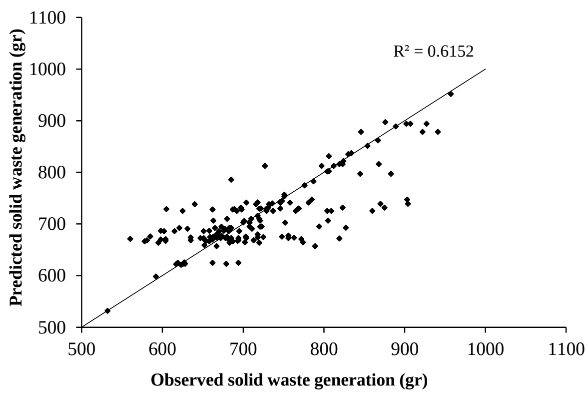
The Cross-validation method with a number of different folds was implemented to validate the M5P model and the model with 13-folds was selected to predict per capita solid waste generation in Hormozgan province due to its lowest prediction error indices. The M5P model with 13 Folds and the root mean square error (RMSE) of 55.34 (gr/d) and the mean absolute relative error (MARE) of 6.26 percent has reached the lowest error. In order to train and validate the SVM and MLP models, 75% and 25% of the dataset were used, respectively. Fig. 1 show the performance of M5P, MLP and SVM model in predicting the per capita waste generation. Also, Table 1. compares the different error evaluation indices of three models in solid waste generation forecasting. According to error indices, M5P model has a better performance in comparison to MLP and SVM.



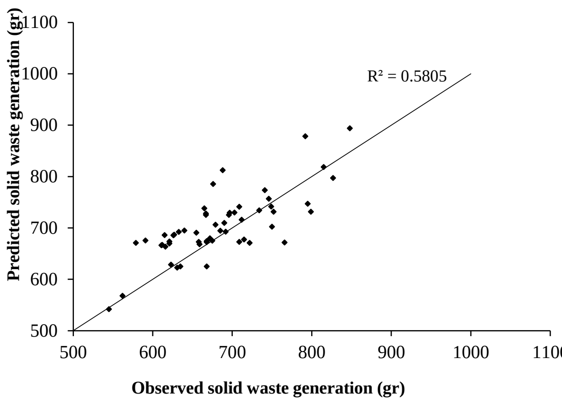
Actual Vs predicted waste generation by MLP (Train data)



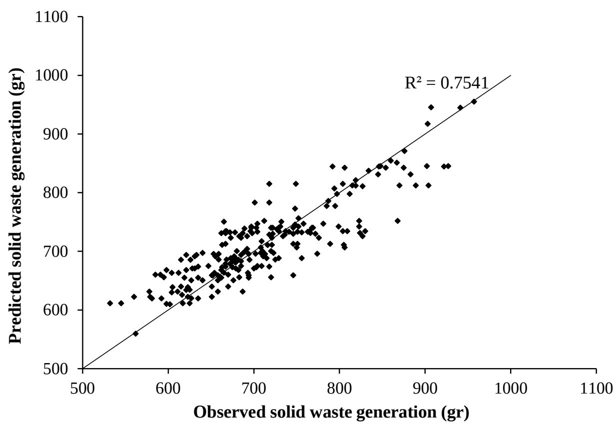
Actual Vs predicted waste generation by MLP (Test data)



Actual Vs predicted waste generation by SVM (Train data)



Actual Vs predicted waste generation by SVM (Test data)



Actual Vs predicted waste generation by M5P

Fig. 1. Actual Vs predicted waste generation by different intelligence techniques

Table 1. Error of the used models to predict the per capita waste production

Models	MARE (%)	RMSE (gr/d)	CC	R^2	MAE (gr/d)
M5P	6.262	55.342	0.87	0.754	44.708
SVM	7.37	66.92	0.78	0.60	52.70
MLP	8.541	70.641	0.79	0.631	58.698

4. Conclusions

According to the results in Table 1, the M5P model with root mean square error (RMSE) (gr/d) 55.34 and the mean absolute relative error (MARE) 6.26 percent has the best performance in comparison with other methods. Furthermore, the correlation coefficient (CC) and R^2 between the results in this case is greater than other methods. Thanks to the simplicity and mathematical relationships between the parameters and the accuracy in predicting solid waste generation, the M5P model can be used as a useful and powerful tool. The comparison between predicted and observed date revealed that using intelligence techniques such as M5P model can be very practical in solid waste generation forecasting especially when there is not enough amount of recorded data for future integrated solid waste management and planning.

5. References

Dyson B, Chang NB, “Forecasting Municipal Solid Waste Generation in a Fast-Growing Urban Region with System Dynamics Modeling”, Waste Management, 2005, Volume 25, Issue 7, 669-679.

Vapnik V, “The Nature of Statistical Learning Theory (Information Science and Statistics)”, New York, NY: Springer-Verlag, 1 edition, January 1995.

Quinlan JR, “Learning with continuous classes”, in Proceedings, AI’92, 5th Australian Joint Conference on Artificial. Intelligence, Adams & Sterling (eds.), World Scientific, Singapore, 1992, 343-348.

Wang Y, Witten I, “Inducing model trees for continuous classes”, In Proceedings of the 9th European Conference on Machine Learning Poster Papers, Prague, 1997, 128-137.