European Journal of Operational Research 250 (2016) 639–651

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

# Hidden Markov model for municipal waste generation forecasting under uncertainties

P. Jiang, X. Liu*

*Department of Industrial Engineering & Management, Shanghai Jiao Tong University, PR China*

ARTICLE INFO

*Article history:*
Received 22 September 2014
Accepted 9 September 2015
Available online 24 September 2015

*Keywords:*
Forecasting
Waste generation
Wavelet de-noising
Hidden Markov model
Gaussian mixture model

ABSTRACT

Waste generation forecasting is a complex process that is found to be influenced by some latent influencing parameters and their uncertainties, such as economic growth, demography, individual behaviors, activities and events, and management policies. These hidden features play an important role in forecasting the fluctuations of waste generation. We therefore focus on revealing the trend of waste generation in megacities which face significant influences of social and economic changes to achieve urban sustainable development. To dynamically trace fluctuations caused by these uncertainties, we propose a probability model-driven statistical learning approach which hybridizes a wavelet de-noising, a Gaussian mixture model, and a hidden Markov model. First, to gain the actual underlying trend, wavelet de-noising is used to eliminate the noise of data. Next, the Expectation–Maximization and the Viterbi algorithms are employed for learning parameters and discerning the most probable sequence of hidden states, respectively. Subsequently, the state transition matrix is updated by fractional predictable changes of influencing parameters to perform non-periodic fluctuation problem forecasting, and the forward algorithm is utilized to search the most similar data pattern for the current pattern from historical data in order to forecast the future trend of the periodic fluctuation problem. Finally, we apply the approaches into two kinds of case studies that test both a small dataset and a large dataset. How uncertainty factors influence forecasted results is analyzed in the subsection of results and discussion. The computational results demonstrate that the proposed approaches are effective in solving the municipal waste generation forecasting problem.



## 1. Introduction

Accurate forecasting of municipal waste generation can provide theoretical guidance in disposal capacity design and capital budget planning for waste management systems (Beigl, Lebersorger, & Salhofer, 2008; Cherian & Jacob, 2012; Intharathirat, Salam, Kumar, & Untong, 2015; Rimaityte, Ruzgas, Denafas, Racys, & Martuzevicius, 2012). Waste generation forecasting in megacities is a complex process that is found to be influenced by some latent influencing factors and their uncertainties, such as rapid economic growth or economic crisis, relentless demographic forces, ever-changing individual behaviors, special activities and events, and complex waste management measures (Cherian & Jacob, 2012; Denafas et al., 2014; Ghiani, Laganà, Manni, Musmanno, & Vigo, 2014; Pires, Martinho, & Chang, 2011). For example, the annual waste generation trend in a megacity could be altered by large-scale activities (such as the Olympic Games and the World EXPO), long-term epidemic diseases (such as SARS), or a new regulation policy for municipal waste. The fluctuation of waste generation comes from unmeasured uncertainties of the economy, individual behaviors, population mobility, and unexpected disasters. The latent factors and their uncertainties are difficult to quantify, and they play an important role in tracing the fluctuation of waste generation dynamically and forecasting the consequences of changes. The fluctuation and nonlinearity of both municipal solid waste (MSW) and wastewater generation is a huge obstacle for accurate forecasting. It is difficult to forecast fluctuation of waste generation accurately via traditional statistical models, such as the moving average and linear regression models.

In this study, we focus on fluctuation forecasting problem. Three major contributions are made in this paper: (1) a probability model-driven statistical learning approach, that hybridizes a wavelet de-noising (WDE), a Gaussian mixture model (GMM), and a hidden Markov model (HMM), is proposed to trace waste generation fluctuations under uncertainties dynamically; (2) the way of weighted average Gaussian mixture and the way of searching the most similar pattern are proposed as two concrete HMM-based forecasting ways to cope with the periodic and non-periodic fluctuation problems, respectively; and (3) two kinds of case studies of different scenarios are presented in this study. They correspond to a small sample and

* Corresponding author at: Department of Industrial Engineering & Management, Shanghai Jiao Tong University, 800 Dongchuan Road, Min-Hang District, Shanghai 200240, PR China. Tel.: +86 65 6601 4089; fax: +86 65 6601 4089.

*E-mail address:* x_liu@sjtu.edu.cn (X. Liu).

http://dx.doi.org/10.1016/j.ejor.2015.09.018
0377-2217/

a large sample datasets to show the effectiveness of the proposed approaches. The first case uses annual MSW generation in Shanghai and Beijing from 1978 to 2010, and the second case utilizes the daily wastewater level of a Spanish urban sewage treatment plant from 1990 to 1991.

The paper is organized as follows: Section 2 summarizes the details of related studies. Problem description and forecasting modelling are presented in Section 3. Section 4 provides two kinds of case studies to test the effectiveness of the proposed approaches. Finally, we close this paper with some conclusions and guidelines for future work.

## 2. Related works: a summary

So far, a number of forecasting techniques have been applied to waste generation forecasting. Mainstream methods include traditional statistical models, the gray and fuzzy models, simulation models, and non-probabilistic statistical learning models. In this section, we summarize the application scope, advantages, and weaknesses of these mainstream methods. Subsequently, literature of probability model-driven statistical learning approaches as a comparison is introduced briefly.

Traditional statistical models include linear regression models (LRM), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), seasonal ARIMA (sARIMA), exponential smoothing (ES), and seasonal ES (sES). LRM aims to discover a linear function to quantify waste generation. Essentially, the problem of waste generation is nonlinear, so LRM is just an approximate fitting method for waste generation forecasting. The ARMA, a mixture of autoregressive and moving average models, can be employed to forecast waste generation of stationary time series (Katsamaki, Willems, & Diamadopoulos, 1998). ARIMA stochastic models are effective time series forecasting tools which can take influence factors into consideration, and can also process periodic time series data effectively. ARIMA model has been applied to a range of forecasting fields, such as portfolio optimization (Ustun & Kasimbeyli, 2012), and MSW generation (Owusu-Sekyere, 2013). The ES model is performed by utilizing the weighted mean of all values in time series, which assigns exponentially decreasing weights over time. Both sARIMA and sES can take into account seasonal changes and trends (Denafas et al., 2014; Rimaityte et al., 2012; Song & He, 2014; Xu, Gao, Cui, & Liu, 2013). Traditional statistical models can forecast the moving average or discover smooth trend effectively; however, fluctuations are hard to be traced accurately.

The gray and fuzzy models are based on fuzzy set theory. The fuzzy mathematic method can solve problems with uncertainty, and can also obtain reliable models given poor data. Chen and Chang (2000) applied a gray fuzzy dynamic prediction technique to forecast MSW generation when the sample numbers were low. Karavezyris, Timpe, and Marzi (2002) employed fuzzy logic to enhance confidence in the validity of the system dynamics model. Fuzzy logic is an attractive method; however, the knowledge base of the fuzzy model is founded on expert experience. Gray model is suitable to study the uncertainty of systems and handle small sample dataset (Intharathirat et al., 2015); however, the conventional gray prediction model is sensitive to initial values (Guo, Liu, Wu, & Tang, 2014).

Simulation models are suitable for simulating complex systems which are difficult to express via mathematic formula. System dynamics (SD) is one of the promising of these methods for studying complex feedback systems. SD models have been widely used for predicting waste generation (Ahmad, 2012; Dyson & Chang, 2005; Karavezyris et al., 2002; Kollikkathara, Feng, & Yu, 2010). SD models use a multivariate method and are far more complex due to the multifarious interactions between the selected parameters (Beigl et al., 2008). Therefore, it is hard to achieve model validation.

Non-probabilistic statistical learning models for waste generation forecasting mainly include artificial neural networks (ANN) and support vector machines (SVM). ANN, one of the most widespread of intelligent statistical-learning models, has powerful ability to deal with nonlinear forecasting problems (Setzler, Saydam, & Park, 2009) and has been applied to generation forecasting fields (Ali Abdoli, Falah Nezhad, Salehi Sede, & Behboudian, 2012; Antanasijevic, Pocajt, Popovic, Redzic, & Ristic, 2013; Jalili Ghazi Zade & Noori, 2008). Although ANN is promising, the disadvantages, such as over-fitting and local minimum, limit its application. SVM is a nonlinear and kernel-based intelligent technique. Abbasi, Abduli, Omidvar, and Baghvand (2013), Noori, Abdoli, Ghasrodashti, and Jalili Ghazizade (2009), and Abbasi, Abduli, Omidvar, and Baghvand (2014)) proposed SVM models to implement weekly prediction of waste generation in the Iranian cities of Mashhad and Tehran, respectively. Compared with ANN, SVM overcomes the problem of over-fitting training data and local minimum. However, SVM is still a black box technique which cannot establish a mathematical functional mapping between the input variables and the drifting parameters (Dong et al., 2009).

Recently, probability model-driven statistical learning approaches have been effectively applied to perform forecasting under uncertainties. Compared with ANN and SVM, they are white box methods which focus more on the components of the mathematical model. For example, Bayesian network (BN) is a combination of graph theory and probability theory which has unique advantages for handling problems with uncertainty, and mining causality among variables. BN has been introduced into forecasting areas, such as traffic flow forecasting (Sun, Zhang, & Yu, 2006) and wastewater inflow monitoring (Cheon, Kim, Kim, & Kim, 2008). The quality of results produced by the BN method largely depends on the amount of evidence (Dong et al., 2009); however, they are not available in most cases. Another example is HMM which became popular for classification and pattern recognition problems after its appearance. The HMM has been applied to a range of fields, such as stock market forecasting (Hassan, 2009), PM2.5 concentration prediction (Dong et al., 2009), failure prognosis (Kim, Jiang, Makis, & Lee, 2011; Zhou, Hu, Xu, Chen, & Zhou, 2010), and manpower planning (Guerry, 2011). However, to the best of our knowledge, no study has previously forecasted waste generation via an HMM.

Although the first four stream methods have respective merits for various applications, there are still some difficulties using these methods for measuring and tracing fluctuations of waste generation dynamically. The advantages of the HMM lie in its theoretical basis, rigorous mathematical structure, and its proven fitness for modelling dynamic systems under uncertainties.

## 3. Model

In this section, the study problems and their characteristics are first introduced. Then, models are constructed to forecast MSW and wastewater generation.

### 3.1. Problem description

Municipal waste generation in megacities is a complex process of the interaction of intertwined systems, which is found to be influenced by latent influencing factors and their uncertainties. So waste generation has the characteristics of nonlinearity, fluctuation, and strong interference. The key challenges for waste generation forecasting problem of both MSW and wastewater lie in generation fluctuations which include non-periodic and periodic fluctuation. Annual MSW generation is a non-stationary time series. The increment rate, transformed from MSW generation, shows an obvious non-periodic fluctuation (Fig. 1). Daily wastewater generation level is an approximate stationary time series with periodic fluctuation (Fig. 2).