2012

Modeling Temporal Pattern and Event Detection using Hidden Markov Model with Application to a Sludge Bulking Data

Naveen K. Bansal
Marquette University, naveen.bansal@marquette.edu

Xin Feng
Marquette University

Wenjing Zhang
Marquette University

Wutao Wi
Marquette University

Yuanhao Zhao
Marquette University

Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Computer Science 12 (2012) 218 – 223
Procedia Computer Science

Complex Adaptive Systems, Publication 2
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2012- Washington D.C.

Modeling temporal pattern and event detection using Hidden Markov Model with application to a sludge bulking data

Naveen K. BANSAL[a], Xin FENG[b], Wenjing ZHANG[b], Wutao WEI[a], and Yuanhao ZHAO[b]

[a]*Department of Math., Stat, and Comp. Sci.,, Marquette University, Milwaukee, Wisconsin, USA*
[b]*Department of Electrical & Computer Engineering, Marquette University, Milwaukee, Wisconsin, USA*

Abstract

This paper discusses a method of modeling temporal pattern and event detection based on Hidden Markov Model (HMM) for a continuous time series data. We also provide methods for checking model adequacy and predicting future events. These methods are applied to a real example of sludge bulking data for detecting sludge bulking for a water plant in Chicago.

Keywords: Temporal Patterns; Hidden Markov Model; Sludge Bulking

1. Introduction

A temporal pattern is a time-ordered fixed structure in a time series data $X = \{x(t) \in \Re \mid t = 1,2,\ldots,n\}$. We consider a situation, in which this temporal pattern occurs repeatedly and may lead to the occurrence of a critical event. The objective is to develop a predictive model for predicting an event based on this temporal pattern. Fig. 1 shows an example of a sample path of a synthetic seismic time series in which pattern occurs somewhat randomly, but that leads to an occurrence of an event. Povinelli and Feng (2003) and Feng and Huang (2005) used a phase space approach to successfully identify temporal pattern and detect event in a non-stationary time series data. This approach, however, is deterministic, and does not reproduce any probabilistic model. In this paper, we use probabilistic hidden Markov model (HMM) to identify pattern and detect events. The details on the model are given in section 2.1. Model estimation and model adequacy checking are described in section 2.2 and 2.3 respectively. In section 2.4 we give a formula for predicting future observations. In section 3, we use a real data on sludge bulking from a water plant in Chicago, US to illustrate statistical methods discussed in section 2.

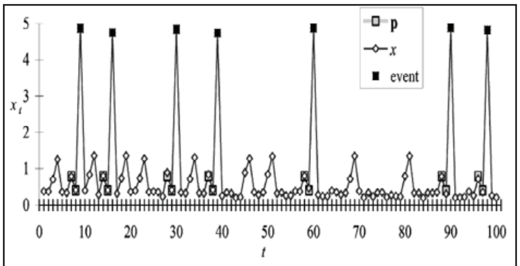

Fig.1 Synthetic seismic time series with temporal pattern **p** and the defined events

2. Modeling Temporal Pattern and Event Occurrence

2.1 Modeling Temporal Pattern and event occurrence

We assume that a time series data $X = \{x(t) \in \Re \mid t = 1,2,\ldots,n\}$ holds a hidden structure of a temporal pattern followed by an event. Suppose that the hidden sequence is $\{Z_t, t = 1,2,\ldots,n\}$, where $Z_t \in S = \{s_0, s_1, \ldots, s_\tau,\ s_{\tau+1}\}$ is a set of all possible states. We call the state s_0 as the normal state, states $\{s_1, s_2, \ldots, s_\tau\}$ as pattern states, and $s_{\tau+1}$ as the event state.

We assume that $\{Z_t, t = 1,2,\ldots\}$ follows a Markov model with the transition probability matrix

doi:10.1016/j.procs.2012.09.059

$$\boldsymbol{P} = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0\tau+1} \\ p_{10} & p_{12} & \cdots & p_{1\tau+1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{\tau+10} & p_{\tau+11} & \cdots & p_{\tau+1\tau+1} \end{bmatrix}, \tag{1}$$

where $p_{ij} = Pr\left(Z_t = s_j \,\middle|\, Z_{t-1} = s_i\right)$.

To structure the temporal pattern and the event occurrence, we partition the state space set S as $S = \{s_0, \boldsymbol{s}_1, s_{\tau+1}\}$, where $\boldsymbol{s}_1 = (s_1, s_2, \dots, s_\tau\}$ is a sequence of the pattern states. In order to reflect the transition of a process from normal to the development of temporal pattern and then to the event, the transition matrix $\boldsymbol{P}$ must take a special form:

$$\boldsymbol{P} = \begin{bmatrix} 1-p & p & 0 & \cdots & 0 \\ 1-q_1 & 0 & q_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1-q_\tau & 0 & \cdots & 0 & q_\tau \\ 1-r & 0 & \cdots & 0 & r \end{bmatrix} \tag{2}$$

Note that $p = P(Z_t = s_1 | Z_{t-1} = s_0)$ is the probability that at a given time t the state would jump from normal to the beginning of the pattern state, but the probability is 0 that it would jump to the second or a higher pattern state. Once the state is in a pattern state, it moves to the next pattern state in a stepwise manner or it goes back to the normal state without completing the pattern. At the last pattern state, when the process completes the pattern, it reaches the event state with probability $q_\tau = P(Z_t = s_{\tau+1} | Z_{t-1} = s_\tau)$. We also assume that once the process reaches the event state, it remains in the event state with probability r or it goes back to the normal state. The main point here is that the event occurs only when the pattern is completed, and if the pattern breaks down, the process goes back to the normal state before the event occurs.

The probability distribution of the observed sequence $\boldsymbol{X}$ is affected by the hidden sequence $\{Z_t, t = 1,2,\dots,n$. Thus, we assume that the observations $Xt, t=1,2,\dots,n$ are conditionally independent given the state sequence $\{Z_t, t = 1,2,\dots,n\}$ with probability model

$$\begin{aligned} &X_t \,|\, Z_t = s_0 \sim f_0(x \,|\, q_0)\,, \\ &X_t \,|\, Z_t = s_1 \sim f_1(x \,|\, \theta_1)\,, \\ &\qquad \vdots \\ &X_t \,|\, Z_t = s_\tau \sim f_\tau(x \,|\, \theta_\tau), \\ &X_t \,|\, Z_t = s_{\tau+1} \sim f_{\tau+1}(x \,|\, \theta_{\tau+1}), \end{aligned} \tag{3}$$

where $f_0, f_1, \dots, f_{\tau+1}$ are the probability densities depending on the unknown parameters $\theta_0, \theta_1, \dots, \theta_{\tau+1}$; for example, f_i can be the Gaussian $N(\mu_i, \sigma_i^2)$ density with $\theta_i = (\mu_i, \sigma_i^2)$. Thus, with little misuse of the notation, the conditional density of $\boldsymbol{X}$ given the hidden states $(z_1, z_2, \dots, z_n)$ can be written as

$$f(x_1, x_2, \dots, x_n | z_1, z_2, \dots, z_n) = f_{z_1}\left(x_1 \middle| \theta_{z_1}\right) f_{z_2}\left(x_2 \middle| \theta_{z_2}\right) \dots f_{z_n}\left(\theta_{z_n}\right),$$

where f_{z_i} and θ_{z_i} are understood to be the appropriate density and parameters respectively corresponding to the actual value of z_i. For the Gaussian model, the assumption of conditional independence can be relaxed by incorporating, for example, ARIMA or GARCH process. This can be done by writing the conditional model for X_t given $Z_t = z_t$ as $X_t = \mu_{z_t} + \sigma_{z_t} \upsilon_t$, where $\{\upsilon_t, t \geq 1\}$ is an ARIMA or GARCH process. In this paper, however, we will only consider the conditional independence case.

2.2 Estimating the Model from a Training Set

Let $(x_1, x_2, \dots, x_n)$ be a training set of data. Baum-Welch (Baum, et al., 1970) method, a method based on the EM algorithm, can be used to estimate the parameters of the model. For illustration purpose, we assume that the conditional densities (3) are the densities of Gaussian $N(\mu_i, \sigma_i^2),\ i = 0,1,\dots,\tau+1$. The unknown parameters of the model are $\boldsymbol{\pi} = (\pi_0, \pi_1, \dots, \pi_\tau, \pi_{\tau+1})$, the probabilities vector of the initial states, the unknown parameters $(p, q_1, \dots, q_\tau, r)$ of the transition matrix, and $\{(\mu_i, \sigma_i^2), i = 0,1,\dots,\tau+1\}$ of the Gaussian model. In most of the cases, the dimension of the pattern states τ may not be known. In such cases, we suggest a phase space embedding method to find the dimension of the pattern states (Albano et al., 1987). Another approach is to use Akaike Information Criteria (AIC), which is defined as

$$AIC = -2\log L + 2\nu, \tag{4}$$

where ν is the total number of parameters. Note that as the dimension of the pattern states increases, the number of parameters ν increases. Thus, the term 2ν on the right hand side of (4) reflects a penalty function of increased dimension.

2.3 Model Adequacy Checking

Note that since $X_t|Z_t = s_j \sim N(\mu_j, \sigma_j^2)$ and since conditionally the observations $X_t, t = 1,2,\dots,n$ are independent, it seems reasonable to consider the standardized residuals

$$R_t = \frac{X_t - \widehat{\mu_J}}{\widehat{\sigma_J}},\ t = 1,2,\dots,n \tag{5}$$

to test the adequacy of the model. To compute the residuals in (5), first the identification of true hidden states for each $i, i = 1,2,\dots,n$ is needed. Viterbi algorithm (Rabiner, 1989) can be used to get the most probable hidden states $\hat{Z}_t = s_{\hat{J}_t},\ t = 1,2,\dots,n$. Thus, an appropriate method of computing the residuals (5) is as follows:

- (i) First apply the Baum-Welch algorithm to get the estimates for $\boldsymbol{\pi} = (\pi_0, \pi_1, \dots, \pi_\tau, \pi_{\tau+1})$, $(p, q_1, \dots, q_\tau, r)$ and $\{(\mu_i, \sigma_i^2), i = 0,1,\dots,\tau+1\}$.
- (ii) Using these estimates, obtain the most probable state $Z_t = s_j$ for each t, $t = 1,2,..n$ by applying Viterbi algorithm.
- (iii) Using the most probable states obtain in (ii), compute the residuals R_t from (5) by substituting the appropriate estimates $(\hat{\mu}_j, \hat{\sigma}_j)$.

These residuals, now, can be used to perform many tasks of residual analysis. Some of the tasks that can be performed are as follows:

- (a) Checking Gaussian distribution: A normal probability plot of the residuals can be used to test the assumption of Gaussian distribution.
- (b) Detecting Outliers: Assuming the Gaussian model, it would be expected that $-3 \le R_t \le 3$. Thus if $|R_t| > 3$ for a t^{th} observation, it can be considered as an outlier. However, note that if $|R_t| > 3$ for too many observations, it would indicate that the proposed Hidden Markov Model is not a good fit.
- (c) Goodness of fit test: A Chi-square goodness of fit statistics $\chi^2 = \sum R_t^2$ can be used to test the goodness of fit. If this $\chi^2 > \chi^2_{n-\nu}(\alpha)$, then one can conclude that the model is not a good fit. Note that here ν is the total number of parameters estimated.
- (d) Serial Correlation: A serial correlation can be tested by looking at the autocorrelations of $R_i s$ or by fitting an ARIMA model to $R_i, i = 1,2,\dots,n$.

2.4 The Prediction

An important problem of fitting a time series with temporal patters and event occurrence is predicting future observations based on the past observed data. Suppose $\{x_t, t = 1,2,\dots.\}$ is the process under consideration, and suppose observations $\boldsymbol{x_t} = (x_1, x_2, \dots, x_t)$ have been observed up to the current time, then the problem of prediction is to predict x_{t+1} based on $\boldsymbol{x_t}$. Note that

$$p(x_{t+1}|\boldsymbol{x_t}) = \sum_j p(x_{t+1}|Z_t = s_j, \boldsymbol{x_t}) p(Z_t = s_j|\boldsymbol{x_t})$$

From this,

$$p(x_{t+1}|\boldsymbol{x_t}) = \sum_j \sum_i p(x_{t+1}|Z_{t+1} = s_i, Z_t = s_j, \boldsymbol{x_t})\, p(Z_{t+1} = s_i|Z_t = s_j, \boldsymbol{x_t}) p(Z_t = s_j|\boldsymbol{x_t})$$

Now, under the assumption of the model, it can be shown that

$$p(x_{t+1}|\boldsymbol{x_t}) = \sum_j \sum_i f_i(x_{t+1}|\theta_i) p_{ji} p(Z_t = s_j|\boldsymbol{x_t}). \tag{6}$$

The future value x_{t+1} can now be estimated by the mode of the posterior distribution (6) or by the posterior mean. The HPD (High posterior density) credible set from (6) would yield most likely values for x_{t+1}. Note that if prediction for a future observation of m time units ahead is required, then it is easy to see that

$$p(x_{t+m}|\boldsymbol{x_t}) = \sum_j \sum_i f_i(x_{t+m}|\theta_i) p_{ji}^m p(Z_t = s_j|\boldsymbol{x_t}), \tag{7}$$

where, p_{ji}^m is the $(j,i)^{th}$ entry of $\boldsymbol{P}^m$ matrix.

If an optimum path of $\{Z_t, t \ge 1\}$ with maximum posterior of $p(Z_t = s_j|\boldsymbol{x_t})$ is available, than a practical alternatives to (6) and (7) are, respectively,

$$p(x_{t+1}|\boldsymbol{x_t}) \approx \sum_i p_{\hat{j}_t i} f_i\left(x_{t+1}\middle|\hat{\theta}_i(t)\right) \tag{8}$$

and

$$p(x_{t+m}|\boldsymbol{x_t}) \approx \sum_i p^m_{\hat{j}_t i} f_i\left(x_{t+1}\middle|\hat{\theta}_i(t)\right) \tag{9}$$

Here, the index $\hat{j}_t$ corresponds to the maximum posterior $p\left(Z_t = s_j\middle|\boldsymbol{x_t}\right)$, and estimate $\hat{\theta}_i(t)$ is based on the estimate from the data $\boldsymbol{x}_t$.

3. Example of Sludge Bulking

In the treatment of sewage, the most commonly used process is an activated sludge process in which air (or pure oxygen) is passed through a mixture of sewage and recycled sludge to allow the micro-organisms to break down the organic components of the sewage. Sludge is continually drawn off as new sewage enters the aeration tank and this sludge must then be settled in sedimentation tanks so that the supernatant can be separated to pass on to further stages of treatment. Sludge bulking occurs when the sludge fails to separate out in the sedimentation tanks, and it is the most notable cause of activated sludge plant failure (i.e. exceeding discharge permit limits) in the U.S. and abroad.

The Sludge Volume Index (SVI) (Dick and Vesilind, 1969) is an empirical measurement for sludge bulking problem. If the sludge bulking occurs, it can generate a high SVI value and very clear supernatant. However, the definition of "High SVI" is different for different wastewater treatment plants (WWTPs) and different research works. Some WWTPs claim that sludge bulking occurs when SVI is larger than 100, while for other plants it is higher than 150, 180, even 200, etc. (Surucu and Soyupak, 1989). A large number of researches have been done to predict high SVI (Sludge Bulking Problem). Some are based on Microscopic Examination Methods, which uses microscope to observe the quantity and categories of the filamentous organisms that causes sludge bulking. Others are based on statistical methods to predict high SVI value. Some of the models considered are Benchmark Simulation Model, AR (Autoregressive) model, ARMA (Autoregressive Moving Average) Model, Risk Assessment Model, etc. (Dennis and Taper, 1994). However, the results have not been very satisfactory.

We propose a hidden-Markov-modeling to model SVI index. It can be postulated that, before the sludge bulking occurs, the process stays in a normal state; and when the sludge bulking is about to occur, the process goes though a sequence of pattern states before the sludge bulking occurs.

3.1. Data and Results

The Metropolitan Water Reclamation District of Greater Chicago (MWRDGC) owns and operates several Water Reclamation Plants (WRPs) in the greater Chicago area and treats roughly 1.4 billion gallons of water daily. Frequently, sludge bulking occurs in the plants. The Fig. 2 shows the daily SVI indices for one of the plants from 2002 to 2006. A significantly high value (greater than 150) in the SVI index is considered to be an indicator of sludge bulking in these plants.

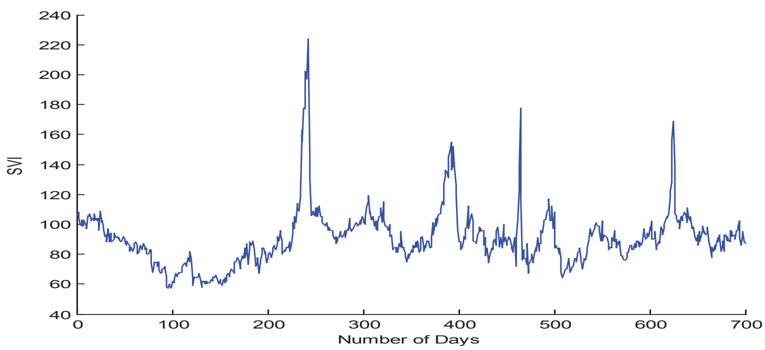

Fig.2. SVI Index Time Series

A data similar to this has been analyzed using traditional methods of regression and time series modeling in the past (Capodaglio et al, 1991). These methods have not been very satisfactory since the data is very chaotic and non-stationary. Here, we use HMM approach to model this data. Because the value of SVI cannot be less than 0, it is incompatible with the Gaussian assumption. We used the logarithm transformation of SVI to achieve the Gaussian distribution.

Using the AIC criteria, we find that the optimal dimension for the pattern states is $\tau = 4$. The initial values of the parameters for the Baum-Welch algorithm were as follows:

$$\boldsymbol{\pi} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \boldsymbol{P} = \begin{bmatrix} .9 & .1 & 0 & 0 & 0 & 0 \\ .1 & 0 & .9 & 0 & 0 & 0 \\ .1 & 0 & 0 & .9 & 0 & 0 \\ .1 & 0 & 0 & 0 & .9 & 0 \\ .1 & 0 & 0 & 0 & 0 & .9 \\ .5 & 0 & 0 & 0 & 0 & .5 \end{bmatrix} \quad \mu = \begin{bmatrix} 80 & 100 & 100 & 100 & 100 & 150 \end{bmatrix}$$

After 68 iterations, the algorithm converged with the following estimates:

$$\hat{\boldsymbol{\pi}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \hat{\boldsymbol{P}} = \begin{bmatrix} .9224 & .0776 & 0 & 0 & 0 & 0 \\ ..0003 & 0 & .9997 & 0 & 0 & 0 \\ .0001 & 0 & 0 & .9999 & 0 & 0 \\ .0009 & 0 & 0 & 0 & .9991 & 0 \\ .8086 & 0 & 0 & 0 & 0 & .1914 \\ ..0910 & 0 & 0 & 0 & 0 & .9090 \end{bmatrix}$$

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} 83.0191 & 99.4780 & 99.3003 & 99.5975 & 100.9446 & 128.9057 \end{bmatrix}$$

In the Fig. 3, the red curves are the patterns we identified, and the red circles are the predicted event points. Note that once the event state is reached, the probability is high (0.9090) that process remains in the event state. On the other hand, the probability is low (0.1914) that the process transit from the last pattern state to the event state meaning that in many occasions before the event (sludge bulking) occurs, the process returns back to the normal state.

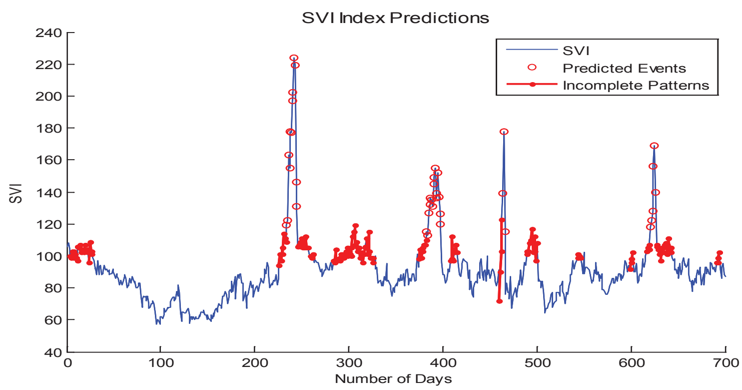

Fig. 3. SVI Index Patterns and Event Predictions

The formula (5) was used to compute the residuals. The normal probability plots of the residuals and the scatter plot of the residuals are shown in the Fig. 4. The probability plot shows slight heavy tail distribution with 6 of the standardized residuals of more than 3 in magnitude. The overall goodness of fit statistic $\chi^2 = \sum R_i^2$ yielding the value of 10.22 with $p - value$ closed to 1, and thus showing a good fit.

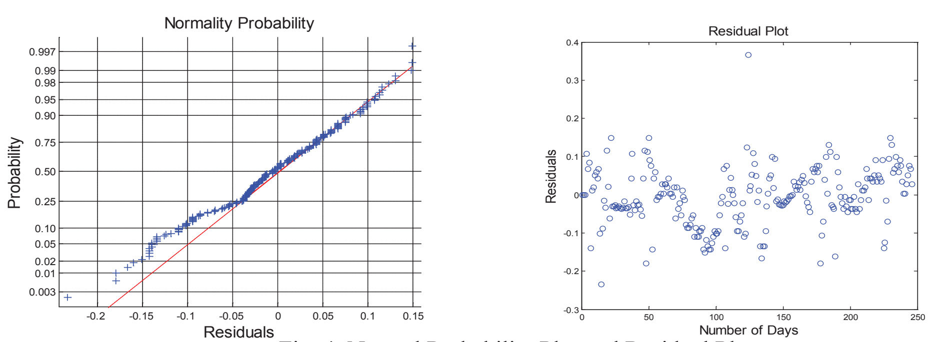

Fig. 4: Normal Probability Plot and Residual Plot

For the prediction purpose, we carry out the analysis with first 700 days' data as training set and the next 250 days' data as test set. The Fig. 5 shows the predictability of the events for the test data. The red circles show the predicted events based on the maximum posterior probabilities. For each prediction, the training set was constantly augmented by new observation x_{t+1} before predicting the next data point x_{t+2} as one would do in the real situation.

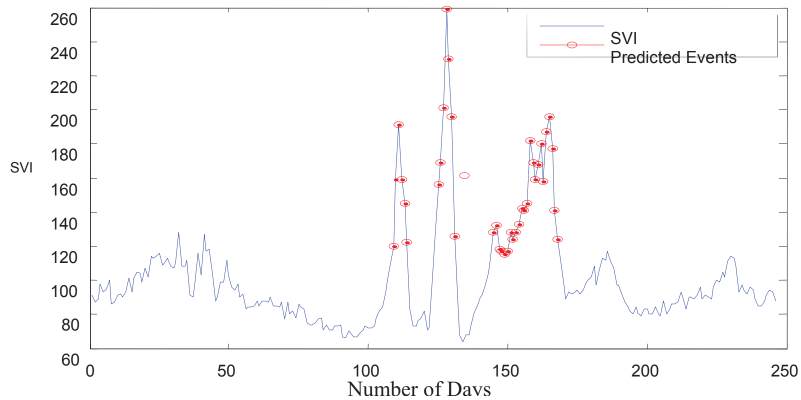

Fig.5. SVI Index Event Predictions

We used equation (8) to approximate the predicting distribution of x_{t+1} based on the previous data $\boldsymbol{x_t} = (x_1, x_2, \ldots, x_t)$. We then computed $P(x_{t+1} > 150|\boldsymbol{x_t})$ which is given by

$$P(x_{t+1} > 150|\boldsymbol{x_t}) \approx \sum_i p_{\hat{j}_t i} \Phi(\frac{150-\hat{\mu}_i}{\hat{\sigma}_i})$$

A higher value of this probability goes along well with the actual SVI index exceeding 150. As we discussed earlier, this has significance in sludge bulking since SVI index of higher than 150 is considered to be an indicator of sludge bulking.

4. Conclusion

When a process is seemingly chaotic and when none of the standard methods such as ARIMA and GARCH models work, the methodology presented in this paper provides an alternative approach through Hidden Markov Modeling. This approach seeks for a problem specific hidden structure in the process. For example, in this paper, we look at the process in which a pattern develops before an event occurs. We provide a complete methodology for this type of problem including model adequacy checking and predicting future values.

References

1. A. M. Albano, A.I. Mees, G.C. de Guzman and P. E. Rapp (1987). "Data Requirements for Reliable Estimation of Correlation Dimensions," *Chaos in Biological Systems,* Plenum Press, New York.
2. L.E. Baum, T. Petrie, G. Soules and N. Weiss (1970). "A maximization technique occurring in the statictical analysis of probabilistic functions of Markov chains," *Ann. Math. Stat*., vol. 41, 164 – 171.
3. A. G. Capodaglio, H. V. Jones, V. Novotny and X. Feng (1991). "Sludge bulking analysis and forecasting: application of system identification and artificial neural computing technologies", *Water Research*; 25(10):1217–24.
4. B. Dennis, and M. L. Taper (1994). Density Dependence in Time Series Observations of Natural Populations: Estimation and Testing. Ecological Monographs 64:205–224.
5. R. I. Dick and P. Aarne Vesilind, "The Sludge Volume Index – What is it?", *Water Pollution Control Federation*, Vol. 41, No. 7 (Jul., 1969) (pp. 1285-1291)
6. X. Feng and H. Huang (2005). "A Fuzzy-Set-Based Reconstruction Phase Space Method for Identification of Temporal Patterns in Complex Time Series," *IEEE Trans. on Knowledge and Data Engineering*, vol.17, no. 5, pp. 601- 613.
7. K. Murphy (2005). "Hidden Markov Model (HMM) Toolbox for Matlab", University of British Columbia.
8. R. J. Povinelli and X. Feng (2003). "A New Temporal Pattern Identification Method for Characterization and Prediction of Complex Time Series Events," *IEEE Trans. on Knowledge and Data Engineering*, vol.15, no. 2, pp.339-352.
9. L. Rabiner (1989). "A tutorial on hidden Markov models and selected applications in speech recognition", *Proceedings of the IEEE*, 257-286.
10. G. Surucu and S. Soyupak (1989). "Effects of operational parameters on the settling properties of activated sludge", *Environmental technology letters*, 471-478.