Prediction and Imputation in Irregularly Sampled Clinical Time Series Data using Hierarchical Linear Dynamical Models

Abhishek Sengupta*[1], Prathosh AP*[2], Satya Narayan Shukla*[3], Vaibhav Rajan*[4], Chandan K Reddy[5]

***Abstract*— Clinical time series, comprising of repeated clinical measurements provide valuable information of the trajectory of patients' condition. Linear dynamical systems (LDS) are used extensively in science and engineering for modeling time series data. The observation and state variables in LDS are assumed to be uniformly sampled in time with a fixed sampling rate. The observation sequence for clinical time series is often irregularly sampled and LDS do not model such data well. In this paper, we develop two LDS–based models for irregularly sampled data. The key idea is to incorporate a temporal difference variable within the state equations of LDS whose parameters are estimated using observed data. Our models are evaluated on prediction and imputation tasks using real irregularly sampled clinical time series data and are found to outperform state-of-the-art techniques.**

I. Introduction

Irregularly sampled time series is commonly found in clinical data. For example, physiological vitals such as blood pressure, heart rate and respiration rate are measured repeatedly but irregularly during a patient's hospital episode. Modeling such time series is important for several applications like clinical decision support systems [1].

Linear dynamical system (LDS) [2], [3], [4] have been extensively used for time series analysis and modelling. They model sequences of measurements (observations) and underlying sequence of states that represent the system dynamics, with the assumption that both the state evolution and measurement sequences are corrupted by noise. Such models attempt to capture the dynamics of the system states that govern the temporal evolution of the measurements. By definition, LDS (described in equation 1), assumes both observation and state variables to be uniformly sampled [2], [5]. Thus, the time difference between any two successive measurement instants is assumed to be a constant, an assumption that is often not true in clinical time series.

In this paper, we design two models, called KF2 and KF3, that explicitly model the time difference between measurements through a temporal difference variable (TDV) and thus can be used for irregularly sampled time series. TDV is additively incorporated into the state evolution equation of LDS in one model (KF2) and two hierarchical LDS are built with the TDV in the another model (KF3). The advantage of our models, compared to previous time series models, are:

- No intermediate transformations are required to model irregularly sampled time series.
- No imputation is required during prediction.
- Our hierarchical model captures the dependency between measurement time intervals and the measurement values, seen in clinical time series as measurements are often more frequent for more severely ill patients.

We evaluate both the models on real patient data comprising irregularly sampled clinical time series (physiological vitals) on prediction and imputation tasks. We show that by extending LDS models to account for the irregularity, we obtain better models of clinical time series. These models when used for prediction and imputation are found to be superior to state-of-the-art methods that assume regular sampling.

A. Related Work

There have been several attempts to address irregular sampling in time series. A popular scheme that is often used is the lifting technique [6], [7] where the observations and state variables are binned into regular periodic intervals in an attempt to convert time-varying multirate systems into a time-invariant single rate system. While this approach is applicable for problems where the sampling pattern is periodic over a larger interval, it becomes intractable when the samples are completely irregularly sampled.

In [5], [8] two approaches are detailed to model irregularly sampled time series. In an approach called the direct value interpolation, a pre-specified fixed sampling rate is assumed and observation values at these points are interpolated using various techniques. Further this interpolated time series is used to train a regular LDS. Another approach is the window-based segmentation approach which is similar to the lifting-based technique. Here the time series is first segmented to intervals of fixed-sized windows. Subsequently, the behavior in each window is summarized in terms of its statistics which are then used in a model such as LDS or a Gaussian process. In both of these approaches, the principle is to convert an irregularly sampled time series into a uniformly sampled time series and subsequently use it in a LDS. Recently, mulitask Gaussian processes (MTGP) have been used [9], [10] to model clinical data where irregular sampling is inherently addressed by using time-dependent kernels which take the instant at which the measurements are made as inputs.

B. Background and Notation

A Linear Dynamical System, also known as Kalman Filter, is given by:

$$\begin{aligned} z_t &= Az_{t-1} + \epsilon_t \\ y_t &= Cz_t + \delta_t \end{aligned} \tag{1}$$

*This research was conducted when these authors were at Xerox Research Centre India, Bangalore, India. [1] Walmart Labs India; [2] Philips Healthcare India; [3] University of Massachusetts Amherst, USA; [4] Yen4Ken Software, India; [5] Virginia Tech, USA.

where, $z_t \in \mathbb{R}^k$ and $y_t \in \mathbb{R}^p$ respectively denote the hidden or state-space variables and observation variables at discrete time intervals $t \in \{1, .., N\}$; $A_{k\times k}$ and $C_{p\times k}$ respectively denote the state-transition and observation matrices; ϵ_t and δ_t are the the noise corrupting the state-space and observation variables, respectively.

In its usual embodiment, a LDS has a deterministic iterative closed-form solution, given the knowledge of system dynamics (A and C) and parameters of the noise involved. The system dynamics are often known from the properties of the physical system being modeled. For example, in a navigation system the physical laws of motion determine the system dynamics. However, when the system dynamics are unknown, the parameters of an LDS can be estimated from (historical/training) data, for example using expectation maximization [11] or spectral estimation methods [12].

We assume a dataset of N patients. For each patient there are p measurement variables. Let $\mathbf{y_i}$ represent the i^{th} patient's multivariate observation sequence of length T_i where $\mathbf{y_i} = \{y^i_{t_1}, ..., y^i_{t_{T_i}}\}$. Note that $\mathbf{y_i} \in \mathbb{R}^{p\times T_i}$ and in general $T_i \neq T_j$ for $i \neq j$. The observation timestamps $t_1, \ldots, t_{T_i}$ differ across patients and are not assumed to be at regular intervals within a patient's observation sequence.

II. Our Models

We introduce a variable called the temporal difference variable (TDV) denoted by Δ defined as follows. For an observation sequence, given a pair of observations instances t_u and t_v, $u \geq v$, Δ is a p-dimensional vector, with the i^{th} dimension, $\Delta^i{}_{t_u,t_v} = t_u - t_v$ if i^{th} data-dimension is observed at both t_u and t_v and zero otherwise. The idea behind the proposed models is to include this variable Δ in the state equation of the LDS so that the state (and thus the observation) at a given time not only depends on the previous state (and observation) but also on the time instant at which the previous observation was made. This information is incorporated in a straightforward manner – shown in model KF2 below. But a better approach is found to be through a hierarchical model – described in model KF3 below.

A. Model KF2

Model KF2 is a direct way of using the temporal difference variable in LDS. To our knowledge, this approach has not been used previously for modeling clinical time series. Δ is linearly scaled by a matrix and added in the state equation:

$$\begin{aligned} z_t &= Az_{t-1} + B\Delta_{t,t-1} + \epsilon_t \\ y_t &= Cz_t + \delta_t \end{aligned} \tag{2}$$

where, $z_t \in \mathbb{R}^k$ denotes the hidden states; $A_{k\times k}$ denotes the state transition matrix; $\Delta_{t,t-1}$ denotes the TDV between the t^{th} and $(t-1)^{th}$ observations; $y_t \in \mathbb{R}^p$ denotes the observations; $C_{p\times k}$ denotes the observation matrix which maps the hidden state space into the observed space. We assume that both the process noise and observation noise are zero-mean Gaussian with an unknown co-variance, that is, $\epsilon_t \sim \mathcal{N}(0, Q)$, $\delta_t \sim \mathcal{N}(0, R)$. It is to be noted that t and $(t-1)$ are used just for notational convenience and they only denote the successive measurement instants in time. Note that in this model, the hidden states at a given time depends on the previous state and also the temporal differences of the measurement instants through B.

Since the noise statistics as well as the system dynamics are unknown for clinical time series, the parameters of the model that are estimated from training data are (A, B, C, Q, R).

B. Model KF3

In KF2, TDV does not directly alter the system dynamics since it does not affect the A matrix. Further, note that the TDV Δ is also a temporal evolving parameter. This is because clinical staff often take more frequent measurements for more severely ill patients. To account for both these observations, we design a hierarchical LDS model where two filters – one on $\Delta_{t,t-1}$ and the other on y – are used.

The filter for $\Delta_{t,t-1}$ can be formulated as:

$$\begin{aligned} \tilde{\Lambda}_{t,t-1} &= \tilde{A}\tilde{\Lambda}_{t-1,t-2} + \tilde{\epsilon}_t \\ \Delta_{t,t-1} &= \tilde{C}\tilde{\Lambda}_{t,t-1} + \tilde{\delta}_t \end{aligned} \tag{3}$$

where, $\tilde{\Lambda}_{t,t-1} \in \mathbb{R}^k$ denotes the hidden states; $\tilde{A}$ denotes the $k \times k$ state transition matrix; $\tilde{\epsilon}_t \sim \mathcal{N}(0, \tilde{Q})$ is the process noise with error covariance $\tilde{Q}$. $\Delta_{t,t-1} \in \mathbb{R}^p$ denotes the observations; $\tilde{C}_{p\times k}$ denotes the observation matrix which maps the true state space into the observed space; $\tilde{\delta}_t \sim \mathcal{N}(0, \tilde{R})$ is the observation noise with error covariance $\tilde{R}$.

The filter for y_t can be formulated as:

$$\begin{aligned} z_t &= A_t z_{t-1} + \epsilon_t \\ y_t &= Cz_t + \delta_t \end{aligned} \tag{4}$$

where, $z_t \in \mathbb{R}^k$ denotes the hidden states; A_t denotes the $k \times k$ state transition matrix at time t; $\epsilon_t \sim \mathcal{N}(0, Q)$ is the process noise with error covariance Q. $y_t \in \mathbb{R}^p$ denotes the observations; $C_{p\times k}$ denotes the observation matrix which maps the true state space into the observed space; $\delta_t \sim \mathcal{N}(0, R)$ is the observation noise with error covariance R.

We model A_t as $A_t = A + B\tilde{\Delta}_{t,t-1}$, where $\tilde{\Delta}_{t,t-1}$ denotes the $k \times k$ **diagonal** matrix with the diagonal entries as $\tilde{\Lambda}_{t,t-1}$, (the hidden states corresponding to equation 3). Note that A_t accumulates the effect of the covariate $\tilde{\Delta}_{t,t-1}$ of the TDV on the final state sequence z_t and is a time-varying matrix. The complete set of parameters for KF3 are $(\tilde{A}, \tilde{C}, \tilde{Q}, \tilde{R}, A, B, C, Q, R)$.

III. Parameter estimation

We briefly sketch the method of parameter estimation for both models KF2, KF3. More details can be found in [11].

For KF2 model described in equation 2, the likelihood of the data (for a single observation sequence) is given by:

$$L(\theta|D) = p(z_1)\prod_{t=2}^{T} p(z_t|z_{t-1})\prod_{t=1}^{T} p(y_t|z_t) \tag{5}$$

$p(z_1) \sim \mathcal{N}(\mu_1, \Sigma_1)$; $p(z_t|zt-1) \sim \mathcal{N}(Az_{t-1} + B\Delta_{t,t-1}, Q)$;$p(y_t|z_t) \sim \mathcal{N}(Cz_t, R)$.

Assuming z_t to be Markovian and from equations 2 and 5, the log likelihood $l(\theta|D)$ for N observation sequences is given by:

$$\begin{aligned} -\frac{1}{2}N\log|\Sigma_1| - \frac{1}{2}\sum_{i=1}^{N} X'\Sigma_1^{-1}X - \sum_{i=1}^{N}\sum_{t=2}^{T_i}\{\frac{1}{2}Y'Q^{-1}Y\} \\ -\frac{1}{2}\sum_{i=1}^{N}(T_i-1)\log|Q| - \sum_{i=1}^{N}\sum_{t=1}^{T_i}\{\frac{1}{2}Z'R^{-1}Z\} \\ -\frac{1}{2}\sum_{i=1}^{N}T_i\log|R| - \frac{1}{2}\sum_{i=1}^{N}T_i(p+k)\log 2\Pi \end{aligned} \tag{6}$$

where $X = (z_{i1} - \mu_1)$ and $Y = (z_{it} - Az_{i(t-1)} - B\Delta_{i,t,t-1})$ and $Z = (y_{it} - Cz_{it})$.

We use Expectation Maximization (EM) to estimate the parameters. Below we state the update equations for i^{th} iteration for the parameters A, B and C below. Other update equations can be derived easily and are not shown.

$$\begin{aligned} A^{(i)} = & (\sum_{i=1}^{N}\sum_{t=2}^{T_i} z_{it}z'_{i(t-1)}\widehat{z_{i(t-1)}} - \\ & \sum_{i=1}^{N}\sum_{t=2}^{T_i} B^{(i-1)}\Delta_{i,t,t-1}z_{i(t-1)}\widehat{z_{i(t-1)}}') \times \\ & (\sum_{i=1}^{N}\sum_{t=2}^{T_i} z_{i(t-1)}z'_{i(t-1)}\widehat{z_{i(t-1)}z'_{i(t-1)}})^{-1} \end{aligned}$$

$$\begin{aligned} B^{(i)} = & (\sum_{i=1}^{N}\sum_{t=2}^{T_i} z_{it}\widehat{z_{it}}\Delta'_{i,t,t-1} - \\ & \sum_{i=1}^{N}\sum_{t=2}^{T_i} A^{(i)}z_{i(t-1)}\widehat{z_{i(t-1)}}\Delta'_{i,t,t-1}) \times \\ & (\sum_{i=1}^{N}\sum_{t=2}^{T_i} \Delta_{i,t,t-1}\Delta'_{i,t,t-1})^{-1} \end{aligned}$$

$$C^{(i)} = (\sum_{i=1}^{N}\sum_{t=1}^{T_i} y_{it}z_{it}\widehat{z_{it}}')(\sum_{i=1}^{N}\sum_{t=1}^{T_i} z_{it}z'_{it}\widehat{z_{it}z'_{it}})^{-1}$$

where $z_{it}\widehat{z_{it}} = E(z_{it}|y_{1:T_i})$; $z_{it}z'_{it}\widehat{z_{it}z'_{it}} = E(z_{it}z'_{it}|y_{1:T_i})$ and $z_{it}z'_{i(t-1)}\widehat{z_{it}z'_{i(t-1)}} = E(z_{it}z'_{i(t-1)}|y_{1:T_i})$ are obtained via Kalman smoothing and filtering equations [13].

In the case of model KF3, there are two filters given by equations 3 and 4. For the first filter (equation 3), parameters are estimated using update equations as described in [11]. For the second filter (equation 4), the log-likelihood is identical to equation 6 except that $Y = (z_{it} - Az_{i(t-1)} - B\tilde{\Delta}_{i,t,t-1}z_{i(t-1)})$. Note that $\tilde{\Delta}_{i,t,t-1}$ is obtained from state equations of the filter from equation 3. Thus the parameters for this filter can be easily obtained using EM, similar to that of KF2.

IV. Experiments and Results

A. Data

To evaluate our models, we use clinical time series of 1000 (randomly sampled) patients from the MIMIC-II publicly available database [14] containing patient records from Intensive Care Units (ICU). We use three vital measurements: Systolic blood pressure (BPS), Respiration Rate (RR) and Heart Rate (HR) for our experiments. The measurement intervals in the data are between 1 second and 1 minute. Measurement intervals outside the ICU is typically much larger, in several hours. To simulate such a setting, we randomly remove measurements to create an irregularly sampled dataset.

B. Experiment Protocol

We test the performance of our model on two tasks: (1) imputation (2) prediction. For the imputation task, in each time series, 20% of the data is randomly removed which then are imputed using our model and other techniques. In the prediction task, given an observation and the instant at which the next observation is to be made (this defines the TDV, Δ), the subsequent observation is predicted. All the experiments are cross-validated over five folds of data splits. The evaluation metrics are the mean and standard deviation of the root mean squared error (RMSE) computed between the imputed/predicted values and the actual values given by $\sqrt{\frac{\sum_{i=1}^{n}(\hat{y_i}-y_i)^2}{n}}$, where $\hat{y_i}$ is the predicted/imputed value given for the true value y_i and n is the total number of predictions made.

C. Baseline methods

For prediction experiments we use three baselines, two of which are previously known: Multi-task Gaussian Process (**MTGP**) [9], [10] and Kalman Filter (**KF1**) (model described in equation 2). Recall that in the KF1 model, we assume (incorrectly) that the sampling is uniform. Thus, given a current observation, the prediction for the next observation is the same irrespective of the instant at which it is made. The third baseline, denoted by **Particle Filter** uses non-linear state-space models (NSS) as described in [15]. Here, the state equation is linear and the observation equations are modeled using a non-linear Gaussian mixed model. However, to handle irregular sampling, we replace the linear state equation by the state equation of KF2 thus using the temporal difference variable. The likelihood of this NSS model is derived using a particle filter [16]. To our knowledge this variant of NSS has not been studied before.

For imputation, We compare our method with state–of–the–art imputation techniques: MICE [17], MTSDI [18] and AMELIA II [19]. MICE and AMELIA are general purpose imputation techniques whereas MTSDI is designed for time series data (implicitly assuming regular sampling).

D. Results

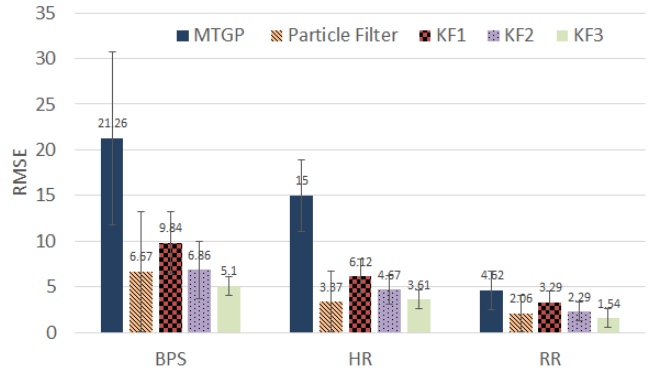

Fig. 1. RMSE for prediction task on Systolic Blood Pressure (BPS), Heart Rate (HR) and Respiration Rate (RR) on patient vitals from MIMIC-II. Lower value is better.

Figure 1 shows the mean RMSE for individual vitals (along with standard deviations) for the prediction task. KF3 outperforms all the other techniques. This implies that using TDV in an LDS hierarchically, models the data very well. It is also seen that the particle filter based model is comparable to KF2 which suggests that a linear model is sufficient to fit the data well. The variance in the performance of KF2 is found to be lesser. The effectiveness of both the models, seen through better performance over KF1, thus can be attributed to the use of TDV. MTGP has lower performance than all the other state-space models.

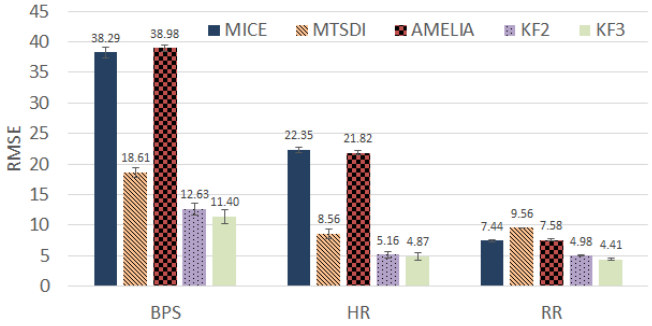

Fig. 2. RMSE for imputation task on Systolic Blood Pressure (BPS), Heart Rate (HR) and Respiration Rate (RR) on patient vitals from MIMIC-II. Lower value is better.

Figure 2 shows the mean RMSE for individual vitals (along with standard deviations) for the imputation task. Among the baselines MTSDI that is designed for time series data outperforms MICE and AMELIA. Both KF2 and KF3 that do not assume regular sampling outperforms MTSDI with KF3 showing the best performance.

V. Conclusion

In this paper, we design models based on linear dynamical systems for irregularly sampled time series data using a temporal difference variable (TDV), that captures the dependencies between consecutive state and observation variables. TDV is additively used in one model (KF2) and hierarchically in the another model (KF3). Prediction and imputation experiments are conducted on irregularly sampled physiological vitals obtained from the MIMIC-II database. Experiments show that our models perform better than state-of-the-art methods thus revealing the value of the TDV in modeling irregular sampling. The dependency of the time interval between measurements on the patient's condition, observed in clinical time series, is also effectively modeled by our hierarchical KF3 model.

References

[1] C. K. Reddy and C. C. Aggarwal, *Healthcare data analytics*. CRC Press, 2015, vol. 36.

[2] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Fluids Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[3] R. H. Shumway and D. S. Stoffer, *Time series analysis and its applications*. Springer Science & Business Media, 2013.

[4] R. E. Kalman, "Mathematical description of linear dynamical systems," *Journal of the Society for Industrial & Applied Mathematics, Series A: Control*, vol. 1, no. 2, pp. 152–192, 1963.

[5] Z. Liu and M. Hauskrecht, "Clinical time series prediction: Toward a hierarchical dynamical system framework," *Artificial intelligence in medicine*, 2014.

[6] J. Sheng, T. Chen, and S. L. Shah, "Generalized predictive control for non-uniformly sampled systems," *Journal of Process Control*, vol. 12, no. 8, pp. 875–885, 2002.

[7] W. Li, S. L. Shah, and D. Xiao, "Kalman filters for non-uniformly sampled multirate systems," in *Proceedings of the 16th Triennial World Congress of the International Federation of Automatic Control*, 2005.

[8] Z. Liu and M. Hauskrecht, "Clinical time series prediction with a hierarchical dynamical system," in *Artificial Intelligence in Medicine*. Springer, 2013, pp. 227–237.

[9] M. Ghassemi, M. A. Pimentel, T. Naumann, T. Brennan, D. A. Clifton, P. Szolovits, and M. Feng, "A multivariate timeseries modeling approach to severity of illness assessment and forecasting in icu with sparse, heterogeneous clinical data," in *Proc. Twenty-Ninth AAAI Conf. on Artificial Intelligence*, 2015.

[10] R. Durichen, M. Pimentel, L. Clifton, A. Schweikard, D. A. Clifton *et al.*, "Multitask gaussian processes for multivariate physiological time-series analysis," *Biomedical Engineering, IEEE Transactions on*, vol. 62, no. 1, pp. 314–322, 2015.

[11] Z. Ghahramani and G. E. Hinton, "Parameter estimation for linear dynamical systems," Technical Report CRG-TR-96-2, University of Totronto, Dept. of Computer Science, Tech. Rep., 1996.

[12] T. Katayama, *Subspace methods for system identification*. Springer Science & Business Media, 2006.

[13] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[14] M. Saeed, C. Lieu, G. Raber, and R. Mark, "Mimic ii: a massive temporal icu patient database to support research in intelligent patient monitoring," in *Computers in Cardiology, 2002*. IEEE, 2002, pp. 641–644.

[15] A. Zia, T. Kirubarajan, J. P. Reilly, D. Yee, K. Punithakumar, and S. Shirani, "An em algorithm for nonlinear state estimation with model uncertainties," *Signal Processing, IEEE Transactions on*, vol. 56, no. 3, pp. 921–936, 2008.

[16] T. B. Schön, A. Wills, and B. Ninness, "System identification of nonlinear state-space models," *Automatica*, vol. 47, no. 1, pp. 39–49, 2011.

[17] S. Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in r," *Journal of statistical software*, vol. 45, no. 3, 2011.

[18] W. Junger and A. P. de Leon, "Imputation of missing data in time series for air pollutants," *Atmospheric Environment*, vol. 102, pp. 96–104, 2015.

[19] J. Honaker, G. King, M. Blackwell *et al.*, "Amelia ii: A program for missing data," *Journal of Statistical Software*, vol. 45, no. 7, pp. 1–47, 2011.