Prediction of Context Information Using Kalman Filter Theory

Mirco Musolesi, Stephen Hailes, Cecilia Mascolo
University College London
Department of Computer Science
Gower Street, London WC1E 6BT, United Kingdom
{m.musolesi|s.hailes|c.mascolo}@cs.ucl.ac.uk

1 Context information prediction using the Kalman filter theory

R.E. Kalman presented in 1960 a novel approach [3] for an efficient solution of the discrete-data linear filtering problem from a computational point of view. The set of recursive equations usually called the Kalman filter has been exploited in a large number of application fields from automatic control systems to weather forecasting.

We apply the Kalman filter theory to the analysis of the time series of values that represent context information. Firstly, we need to have a representation of the problem according a typical state space model, since we have a set of observations and we derive a prediction model based on an inner state that is represented by a set of vectors. In the following section, we give a general introduction to state space models and, in the remainder of this appendix, we present the ways in which we have applied these concepts to the analysis and the prediction of context information, discussing three cases according to the different behaviour of the time series. More specifically, we consider the cases of time series characterised by local trends and seasonal trends [1]. Clearly, the complete model introduces more computational overhead; however, it provides more accurate results.

1.1 State space models

A state space model for a time series $\mathbf{Y}_t$ consists of two equations. The first one called the *observation equation* is the following

$$\mathbf{Y}_t = G_t \mathbf{X}_t + \mathbf{W}_t \quad t = 1, 2, ...$$

with $\mathbf{W}_t$ defined as [1]

$$\mathbf{W}_t = WN(0, R_t)$$

This equation defines the w-dimensional observation $\{\mathbf{Y}_t\}$ as a linear function of a v-dimensional state variables $\{\mathbf{X}_t\}$ and a noise term. The second one is the *state equation* defined as follows

$$\mathbf{X}_{t+1} = F_t \mathbf{X}_t + \mathbf{V}_t \quad t = 1, 2, ...$$

with $\mathbf{V}_t$ defined as

$$\mathbf{V}_t = WN(0, Q_t)$$

This equation determines the state $\mathbf{X}_{t+1}$ at time $t+1$ in terms of the previous state $\mathbf{X}_t$ and a noise term. Defining w as the dimension of $\mathbf{Y}_t$ and v as the dimension of $\mathbf{X}_t$, $\{G_t\}$ is a sequence of $w \times v$ matrices and $\{F_t\}$ is a sequence of $v \times v$ matrices. We assume that $\{\mathbf{V}_t\}$ is uncorrelated with $\{\mathbf{W}_t\}$, even if a more general form of the state space model allows for correlation between these two variables. Analytically, we can rewrite this condition as follows

$$E(\mathbf{W}_t, \mathbf{V}_t^T) = 0 \quad \forall\, s, t$$

We also assume that the initial state $\mathbf{X}_1$ is uncorrelated with all of the noise terms $\{\mathbf{V}_t\}$ and $\{\mathbf{W}_t\}$.

1.2 Kalman filter prediction

With the notation of $P_t(\mathbf{X})$ we refer to the best linear predictor (in the sense of minimum mean-square error) of $\mathbf{X}$ in terms of $\mathbf{Y}$ at the time t. $P_t(\mathbf{X})$ is defined as follows

$$P_t(\mathbf{X}) \equiv \left[\begin{array}{ccc} P_t(X_1) & ... & P_t(X_v) \end{array} \right]^T$$

where

$$P_t(X_i) \equiv P(X_i | \mathbf{Y}_0, \mathbf{Y}_1, ..., \mathbf{Y}_t)$$

$P(X_i | \mathbf{Y}_0, \mathbf{Y}_1, ..., \mathbf{Y}_t)$ indicates the best predictor of X_i given $\mathbf{Y}_0, ..., \mathbf{Y}_t$. We can also observe that $P_t(\mathbf{X})$ has the following form

$$P_t(\mathbf{X}) = A_0 \mathbf{Y}_0 + ... + A_t \mathbf{Y}_t$$

since it is a linear function of $\mathbf{Y}_0, ..., \mathbf{Y}_t$. It is possible to prove [1] for the state space model discussed in the previous Section that the one-step predictor

$$\widehat{\mathbf{X}}_t \equiv P_{t-1}(\mathbf{X}_t)$$

and their error covariance matrices

$$\Omega_t = E[(\mathbf{X}_t - \widehat{\mathbf{X}}_t)(\mathbf{X}_t - \widehat{\mathbf{X}}_t)^T]$$

[1] WN stands for White Noise, a term that derives from telecommunication engineering. A white noise is a sequence of uncorrelated random variables X_t, each with the same mean and variance σ^2. Therefore, white noise is also an example of stationary time series. More specifically, the notation $WN(0, \{R_t\})$ indicates white noise with zero mean and variance R_t.

are determined by these initial conditions

$$\widehat{\mathbf{X}}_1 = P(\mathbf{X}_1|\mathbf{Y}_0)$$

$$\Omega_1 = E[(\mathbf{X}_1 - \widehat{\mathbf{X}}_1)(\mathbf{X}_1 - \widehat{\mathbf{X}}_1)^T]$$

and these recursive equations

$$\widehat{\mathbf{X}}_{t+1} = F_t\widehat{\mathbf{X}}_t + \Theta_t\Delta_t^{-1}(\mathbf{Y}_t - G_t\widehat{\mathbf{X}}_t)$$

$$\Omega_{t+1} = F_t\Omega_t F_t^T + Q_t - \Theta_t\Delta_t^{-1}\Theta_t^T$$

where

$$\Delta_t = G_t\Omega_t G_t^T + R_t$$
$$\Theta_t = F_t\Omega_t G_t^T$$

1.3 Estimation models

1.3.1 Basic model

The basic state space model is composed of the following two scalar equations

$$Y_t = X_t + W_t \quad t = 1, 2, ...$$

with

$$W_t = WN(0, Q_t)$$

and

$$X_{t+1} = X_t + V_t \quad t = 1, 2, ...$$

with

$$V_t = WN(0, R_t)$$

With respect to the Kalman filter prediction we can consider a mono-dimensional system with

$$G_t = [1]$$

$$F_t = [1]$$

Therefore, we can derive the recursive equations of the Kalman filter for the prediction of the values of this series. Given the previous observed value Y_t and the predicted value at time t, $\widehat{X}_t$, the recursive equation for the determination of the predicted value at time $t+1$ is

$$\widehat{X}_{t+1} = \widehat{X}_t + \frac{\Omega_t}{\Omega_t + R_t}(Y_t - \widehat{X}_t)$$

with

$$\Omega_{t+1} = \Omega_t + Q_t - \frac{\Theta_t^2}{\Omega_t + R_t}$$

Since in this case

$$\Omega_t = \Theta_t$$

we can also write

$$\Omega_{t+1} = \Omega_t + Q_t - \frac{\Omega_t^2}{\Omega_t + R_t}$$

1.3.2 Model with trend component

A more complex model can be obtained adding a trend component. We adopt a model composed by the following equations:

$$\begin{cases} Y_t = M_t + W_t \\ M_{t+1} = M_t + B_t + V_t \\ B_{t+1} = B_t + U_t \end{cases}$$

where

$$W_t = WN(0, \sigma_w^2)$$

$$V_t = WN(0, \sigma_v^2)$$

$$U_t = WN(0, \sigma_u^2)$$

In this model the state vector is the following

$$\mathbf{X}_t = \begin{bmatrix} M_t \\ B_t \end{bmatrix}$$

We can write

$$\begin{bmatrix} M_{t+1} \\ B_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} M_t \\ B_t \end{bmatrix} + \begin{bmatrix} V_t \\ U_t \end{bmatrix}$$

Setting

$$\mathbf{V}_t = \begin{bmatrix} V_t \\ U_t \end{bmatrix}$$

we can also rewrite this equation in a more compact way

$$\widehat{\mathbf{X}}_{t+1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \widehat{\mathbf{X}}_t + \mathbf{V}_t$$

Using the same notation that we have adopted for the Kalman filter, in this case we have

$$F_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$G_t = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

$$Q_t = \begin{bmatrix} \sigma_V^2 & 0 \\ 0 & \sigma_U^2 \end{bmatrix}$$

$$R_t = \sigma_W^2$$

Therefore, we can rewrite the Kalman filter prediction equations for this model. Firstly, we consider the initial conditions that, in this case, can be calculated using the following formulae

$$\widehat{\mathbf{X}}_1 = P(\mathbf{X}_1|Y_0)$$

$$\begin{aligned} \Omega_1 &= E((\mathbf{X}_1 - \widehat{\mathbf{X}}_1)(\mathbf{X}_1 - \widehat{\mathbf{X}}_1)^T) = \\ &= E\left(\begin{bmatrix} M_1 - \widehat{M}_1 \\ B_1 - \widehat{B}_1 \end{bmatrix} \begin{bmatrix} M_1 - \widehat{M}_1 & B_1 - \widehat{B}_1 \end{bmatrix} \right) = \\ &= E\begin{pmatrix} (M_1 - \widehat{M}_1)(M_1 - \widehat{M}_1) & (M_1 - \widehat{M}_1)(B_1 - \widehat{B}_1) \\ (M_1 - \widehat{M}_1)(B_1 - \widehat{B}_1) & (B_1 - \widehat{B}_1)(B_1 - \widehat{B}_1) \end{pmatrix} \end{aligned}$$

With respect to the recursive equations of the filter we obtain

$$\widehat{\mathbf{X}}_{t+1} = \begin{bmatrix} \widehat{M}_{t+1} \\ \widehat{B}_{t+1} \end{bmatrix} =$$
$$= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \widehat{M}_t \\ \widehat{B}_t \end{bmatrix} + \Delta_t^{-1}(Y_t - M_t) \begin{bmatrix} \Theta_{M_t} \\ \Theta_{B_t} \end{bmatrix} =$$
$$= \begin{bmatrix} \widehat{M}_t + \widehat{B}_t \\ \widehat{B}_t \end{bmatrix} + \Delta_t^{-1}(Y_t - M_t) \begin{bmatrix} \Theta_{M_t} \\ \Theta_{B_t} \end{bmatrix}$$

that can be decomposed as follows

$$\widehat{M}_{t+1} = \widehat{M}_t + \widehat{B}_t + \Delta_t^{-1}(Y_t - M_t)\Theta_{M_t}$$
$$\widehat{B}_{t+1} = \widehat{B}_t + \Delta_t^{-1}(Y_t - B_t)\Theta_{B_t}$$

with

$$\Delta_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \left[\sigma_w^2\right] = \left[\Omega_{11}\right] + \left[\sigma_w^2\right]$$

1.3.3 Model with trend and seasonal components

It is possible to derive a more general model, adding another component in order to allow consideration of possible seasonal behaviour in the time series. Therefore, we introduce the term S_t in the observation equation, which can be rewritten as follows

$$Y_t = M_t + B_t + S_t + W_t$$

To define this seasonal component we have to analyse its properties. In general, we consider a time series γ_t representing a seasonal component such that

$$\gamma_{t+d} = \gamma_t$$

and

$$\sum_{i=1}^{d} \gamma_t = 0$$

Therefore, it is possible to derive the following expression for the determination of γ_{t+1}

$$\gamma_{t+1} = -\gamma_t - ... - \gamma_{t-d+2} \qquad t = 1, 2, ...$$

A more general expression of the seasonal component S_t allowing for random deviations from strict periodicity is obtained by adding a term V_t to the right hand side of the previous expression

$$S_{t+1} = -S_t - ... - S_{t-d+2} + V_t \qquad t = 1, 2, ..$$

Considering only the seasonal effect, in order to obtain a state space representation, we introduce the $(d-1)$-dimensional state vector $\mathbf{X}_t$

$$\mathbf{X}_t = \begin{bmatrix} S_t & S_{t-1} & ... & S_{t-d+2} \end{bmatrix}^T$$

The series S_t is given by the observation equation

$$S_t = \begin{bmatrix} 1 & 0 & 0 & 0 & ... & 0 \end{bmatrix} \mathbf{X}_t \qquad t = 1, 2, ...$$

where $\mathbf{X}_t$ satisfies the state equation

$$\mathbf{X}_{t+1} = F\mathbf{X}_t + \mathbf{V}_t \qquad t = 1, 2, ...$$

with

$$\mathbf{V}_t = \begin{bmatrix} Z_t & 0 & 0 & 0 & ... & 0 \end{bmatrix}^T$$

and

$$F = \begin{bmatrix} -1 & -1 & \cdots & -1 & -1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

To derive a prediction state space model with a trend and a seasonal components, it is sufficient to add this state equation to that which we have discussed in the previous section. In other words, we have to consider the following state vectors

$$\mathbf{X}_t^1 = \begin{bmatrix} M_t & B_t \end{bmatrix}^T$$
$$\mathbf{X}_t^2 = \begin{bmatrix} S_t & S_{t-1} & ... & S_{t-d+2} \end{bmatrix}^T$$

Defining

$$\mathbf{X}_t = \begin{bmatrix} X_t^1 \\ X_t^2 \end{bmatrix}$$

we can derive a general form of the state equation that can be used to take in consideration both trend and seasonal components as follows

$$\mathbf{X}_t = F\mathbf{X}_t + \mathbf{V}_t$$

with

$$\mathbf{X}_t = \begin{bmatrix} M_t & B_t & S_t & S_{t-1} & ... & S_{t-d+2} \end{bmatrix}^T$$

$$F = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & -1 & \cdots & -1 & -1 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

$$\mathbf{V}_t = \begin{bmatrix} V_t & U_t & Z_t & 0 & ... & 0 \end{bmatrix}^T$$

The observation equation will be the following

$$Y_t = \begin{bmatrix} 1 & 0 & 1 & 0 & ... & 0 \end{bmatrix} \mathbf{X}_t + W_t$$

1.4 Subsequent predictions

It is also possible to make subsequent predictions, in the case that observation values are not available, by using the same model that we discussed previously [1, 2]. In other words, it is possible to predict h consecutive estimations $\mathbf{Y}_{t+1}, ..., \mathbf{Y}_{t+h}$, using the predictor calculated at time without exchanging new context and routing information[2]. Now we derive the expression of the h-step prediction of $\mathbf{Y}_t$ using recursive equations.

[2] It is possible to evaluate the maximum number n of predictions that it is possible to make with a given required accuracy. However, we do not present these results in detail, since they are outside the scope of this work.

Observing that

$$\widehat{\mathbf{X}}_t = P_t(\mathbf{X}_{t+1}) = F_t P_{t-1}(\mathbf{X}_t) + \Theta_t \Delta_t^{-1}(\mathbf{Y}_t - P_{t-1}(\mathbf{X}_t))$$

it is possible to write $\widehat{\mathbf{X}}_{t+h}$ as

$$\begin{aligned}
&\widehat{\mathbf{X}}_{t+h} = P_t(\mathbf{X}_{t+h}) = \\
&= F_{t+h-1} P_t(\mathbf{X}_{t+h-1}) = \\
&\vdots \\
&= (F_{t+h-1} F_{t+h-2} ... F_{t+1}) P_t(\mathbf{X}_{t+1}) \qquad t = 2, 3, ...
\end{aligned}$$

and $\widehat{\mathbf{Y}}_{t+h}$ as

$$\widehat{\mathbf{Y}}_{t+h} = P_t(\mathbf{Y}_{t+h}) = G_{t+h} P_t(\mathbf{X}_{t+h})$$

References

[1] Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting*. Springer, 1996.

[2] James Durbin and Siem Jan Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press, 2001.

[3] Rudolph. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME Journal of Basic Engineering*, March 1960.