Multivariate Forecasting of Operational Times Using Hidden Markov Models

Fernando Miguelez[12], Josu Doncel[3], Maria Dolores Ugarte[12]

[1] Department of Statistics, Computer Science and Mathematics, Public University of Navarre, Spain
[2] Institute for Advanced Materials and Mathematics (InaMat[2]), Public University of Navarre, Spain
[3] Department of Mathematics, University of the Basque Country, Spain

E-mail for correspondence: `fernando.miguelez@unavarra.es`

Abstract: Industrial processes generate a massive amount of monitoring data that can be exploited to uncover hidden time losses in the system, leading to enhanced accuracy of maintenance policies and, consequently, increasing the effectiveness of the equipment. In this work, we propose a method for one-step probabilistic multivariate forecasting of time variables based on a Hidden Markov Model with covariates (IO-HMM). These covariates account for the correlation of the predicted variables with their past values and additional process measurements by means of a discrete model and a continuous model. The probabilities of the former are updated using Bayesian principles, while the parameter estimates for the latter are recursively computed through an adaptive algorithm that also admits a Bayesian interpretation. This approach permits the integration of new samples into the estimation of unknown parameters, computationally improving the efficiency of the process. We evaluate the performance of the method using a real data set obtained from a company in the food sector; however, it is a versatile technique applicable to any other data set. The results show a consistent improvement over a persistence model, which assumes that future values are the same as current values, and over univariate versions of our model.

Keywords: Adaptive parameter estimates; Hidden Markov Model; Industrial processes; Probabilistic prediction

1 Background and objectives

In industrial settings, production processes often face inefficiencies that lead to time losses. These time losses can be broadly classified into four

This paper was published as a part of the proceedings of the 38th International Workshop on Statistical Modelling (IWSM), Durham, UK, 14–19 July 2024.

categories (Muchiri and Pintelon, 2008): losses due to scheduled stops such as maintenance or cleaning; losses due to unexpected stops such as setup, adjustment, failure, or supply outage; losses due to low production speed and micro-stoppages; and losses due to the production of defective units and rework. One can also derive different production times by successively subtracting each time loss from the total length of the observation period, as well as some important efficiency indexes as ratios of these production times.

In this work, we propose a novel approach to predict time losses by modelling the production process carried out by the equipment as a multi-signal process, where the signals characterize the equipment's current operational mode. Furthermore, the predictive model includes other process features that can have an impact on the model parameters as covariates. To ensure continuous parameter updating using the latest data, we use an adaptive learning algorithm that admits a Bayesian interpretation. The forecasting of time losses in production processes can help to enhance the maintenance strategy's accuracy by identifying areas for improvement.

2 The model

We use an Input-Output Hidden Markov Model (IO-HMM) to model the production process, see (Bengio and Frasconi, 1996) for full details of IO-HMMs. Figure **??** illustrates an IO-HMM diagram. The process goes through K hidden states according to an initial state probability distribution and a transition probability distribution between states. The hidden state of the n-th observation period is denoted by c_n and represents the condition of the production process during that period. Each state gives rise to a different probability distribution of the continuous responses $\mathbf{y}_n$.

In an IO-HMM, the model's probability distributions are affected by an input stream of covariates, denoted by $\mathbf{x}_n$. These covariates may include, among others, calendar variables or the reference produced, and characterize the observation period that is about to begin. Further, we introduce an autoregressive component into the model by allowing the covariates to include past values of the response variables. The covariates that influence the probabilities in the discrete part of the model will be denoted by $\mathbf{z}_n \subseteq \mathbf{x}_n$, while the ones that impact the responses' joint density will be denoted by $\mathbf{w}_n \subseteq \mathbf{x}_n$.

3 Parameter Estimates

We consider that the discrete process $\{c_n\}_{n\geq 1}$ is a Markov chain with K different states, $c_n \in \{1,\ldots,K\}\,,\ n \geq 1$. The probability distributions for the initial state and the transitions between states are dependent on the covariates $\mathbf{z}_n$, which take values in a discrete and finite set of S symbols.

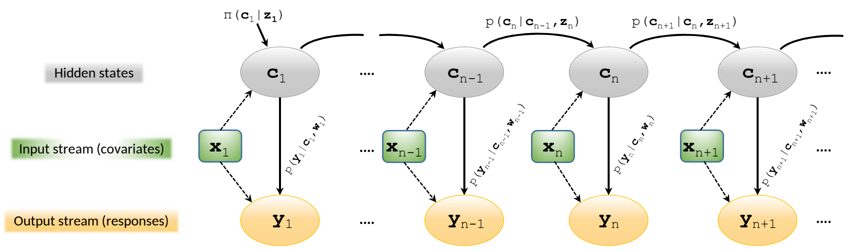

FIGURE 1. Input-Output HMM. Covariates $\mathbf{x}_n$ affect both discrete and continuous processes. Probabilities in the discrete process $\{c_n\}_{n\geq 1}$ are dependent on covariates $\mathbf{z}_n \subseteq \mathbf{x}_n$ and probabilities in the continuous process $\{\mathbf{y}_n\}_{n\geq 1}$ are dependent on covariates $\mathbf{w}_n \subseteq \mathbf{x}_n$.

The unobserved next state c_n is categorical with parameters $\pi^{(s)}$ if $n = 1$ and $\mathbf{z}_n = s$, or $\mathbf{p}_k^{(s)}$ when $n > 1$ and $\mathbf{z}_n = s$. In turn, $\pi^{(s)}$ and $\mathbf{p}_k^{(s)}$ are Dirichlet with parameters - i.e., counts - updated every time an observation period ends.
On the other hand, we split the responses' joint density function into two conditional Gaussian distributions, namely

$$\mathbf{y}_n|\mathbf{w}_n \sim \mathcal{N}_m(\mathbf{u}_n\mathbf{H}_u, \mathbf{\Sigma}_u) \tag{1}$$
$$\mathbf{y}_n|c_n \sim \mathcal{N}_m(\mathbf{v}_n\mathbf{H}_v, \mathbf{\Sigma}_v), \tag{2}$$

where $\mathbf{u}_n = \left[1\ \mathbf{w}_n^T\right]$, $\mathbf{v}_n = v(c_n)$ for a function $v(\cdot)$, $\mathbf{H}_u, \mathbf{H}_v$ are coefficient matrices and $\mathbf{\Sigma}_u, \mathbf{\Sigma}_v$ are covariance matrices. As soon as a new sample $\mathbf{y}_n$ becomes available, the estimators $(\mathbf{H}_{u,n-1}, \mathbf{\Sigma}_{u,n-1}, \mathbf{H}_{v,n-1}, \mathbf{\Sigma}_{v,n-1})$ are updated to $(\mathbf{H}_{u,n}, \mathbf{\Sigma}_{u,n}, \mathbf{H}_{v,n}, \mathbf{\Sigma}_{v,n})$ through an adaptive algorithm described by the multivariate extension of the equations introduced by Alvarez et al. (2021)

$$\mathbf{H}_{u,n} = \mathbf{H}_{u,n-1} + \frac{\mathbf{P}_{u,n-1}\mathbf{u}_n^T}{\lambda_u + \mathbf{u}_n\mathbf{P}_{u,n-1}\mathbf{u}_n^T}\left(\mathbf{y}_n - \mathbf{u}_n\mathbf{H}_{u,n-1}\right)$$
$$\mathbf{\Sigma}_{u,n} = \mathbf{\Sigma}_{u,n-1} - \frac{1}{\gamma_{u,n}}\left[\mathbf{\Sigma}_{u,n-1} - \frac{\lambda\left(\mathbf{y}_n - \mathbf{u}_n\mathbf{H}_{u,n-1}\right)^T\left(\mathbf{y}_n - \mathbf{u}_n\mathbf{H}_{u,n-1}\right)}{\lambda + \mathbf{u}_n\mathbf{P}_{u,n-1}\mathbf{u}_n^T}\right]$$
$$\mathbf{P}_{u,n} = \frac{1}{\lambda_u}\left(\mathbf{P}_{u,n-1} - \frac{\mathbf{P}_{u,n-1}\mathbf{u}_n^T\mathbf{u}_n\mathbf{P}_{u,n-1}}{\lambda_u + \mathbf{u}_n\mathbf{P}_{u,n-1}\mathbf{u},_n^T}\right)$$
$$\gamma_{u,n} = 1 + \lambda_u\gamma_{u,n-1},$$

where λ_u is a forgetting factor. The algorithm is initialized with $\mathbf{H}_{u,0} = \mathbf{0}$, $\mathbf{\Sigma}_{u,0} = \mathbf{0}$, $\mathbf{P}_{u,0} = \mathbf{I}$ and $\gamma_{u,0} = 0$. The same updating equations are applied to compute $\mathbf{H}_{v,n}$ and $\mathbf{\Sigma}_{v,n}$ with the vector $\mathbf{v}_n$ and the forgetting factor λ_v.

4 Forecasting

At this stage each distribution produces a forecast of the responses, which are then combined using a minimum-variance criterion to obtain the final prediction. In particular, once the parameters are updated at the n-th time step the model computes the final prediction and a measure of its accuracy as

$$\hat{\mathbf{y}}_{n+1} = \mathbf{u}_{n+1}\mathbf{H}_{u,n}\mathbf{D} + \mathbf{v}_{n+1}\mathbf{H}_{v,n}(\mathbf{I} - \mathbf{D})$$
$$\hat{\mathbf{\Sigma}}_{n+1} = \mathbf{D}\mathbf{\Sigma}_{u,n}\mathbf{D} + (\mathbf{I} - \mathbf{D})\mathbf{\Sigma}_{v,n}(\mathbf{I} - \mathbf{D}),$$

where $\mathbf{D} = \text{diag}\,(\delta_1, \ldots, \delta_m)$, $\delta_j = \sigma^2_{v,j} / \left(\sigma^2_{v,j} + \sigma^2_{v,j}\right)$, $j = 1, \ldots, m$, and $\sigma^2_{v,j}$ (respectively $\sigma^2_{u,j}$) is the j-th element in the diagonal of $\mathbf{\Sigma}_{v,n}$ (respectively $\mathbf{\Sigma}_{u,n}$).

5 Real case study

The proposed model has been employed to predict time losses in the production process of a company that operates in the food industry. To measure the predictions' quality we use the well-known metrics Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The multivariate predictive model with an autoregressive component in the covariates $\mathbf{w}_n$ shows a consistent improvement in the predictions' quality against some benchmark models, including the persistence model, which assumes that future values are the same as current values (i.e., $\hat{\mathbf{y}}_{n+1} = \mathbf{y}_n$), the model with no autoregressive component and the respective univariate versions of our model.

Acknowledgments: This research was supported in part by the Government of Navarre under Smart Zeo Project 0011-1365-2021-000085.

References

Alvarez, V., Mazuelas, S. and Lozano, J.A. (2021). Probabilistic load forecasting based on adaptive online learning. *IEEE Transactions on Power Systems*, **36**, 4, 3668 – 3680.

Bengio, Y. and Frasconi, P. (1996). Input-Ouput HMMs for sequence processing. *IEEE Transactions on Neural Networks*, **7**, 5, 1231 – 1249.

Muchiri, P. and Pintelon, L. (2008). Performance measurement using overall equipment effectiveness (OEE): literature review and practical application discussion. *International Journal of Production Research*, **46**, 13, 3517 – 3535.