ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

A hidden Markov model with Hawkes process-derived contextual variables to improve time series prediction. Case study in medical simulation.

Fatoumata Dama[1], Christine Sinoquet[1] and Corinne Lejus-Bourdeau[2] *

1- LS2N, UMR CNRS 6004, Nantes Université
2 rue de la Houssinière, 44322 Nantes Cedex, France

2 - CHU de Nantes, Nantes Université
1 place Alexis-Ricordeau, 44000 Nantes, France

Abstract. So far, models that take advantage of sequences of events to refine time series prediction have only been designed for specific applications. In this paper, we introduce the Non-Homogeneous Markov Chain AutoRegressive (NHMC-AR) model. In our model, the innovation arises from the synchronization of a multivariate Hawkes temporal point process with an autoregressive first-order hidden Markov model, through contextual variables. Experiments on anaesthesia data demonstrate that NHMC-AR has substantially better predictive performance compared to two competing methods.

1 Introduction and motivation

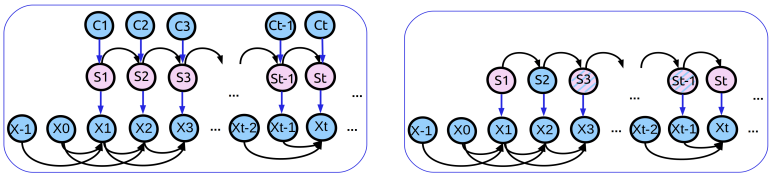
The histories of events in systems or their environments are now being collected alongside multivariate time series (MTSs) as event logs (ELs), in an increasing number of systems. We therefore face unprecedented opportunities to develop frameworks for jointly modelling an MTS and an EL. Such joint models are largely absent to date, except for specific tasks such as survival analysis and rare event prediction [1].

The motivation for the joint model presented here originates from a demand of healthcare professionals for data-driven simulation training in anaesthesia. Since 2000, Nantes University Hospital has been recording anaesthesia profiles for all surgical procedures performed. An anaesthesia profile consists in an MTS and an EL, for a patient characterized by sex, age, weight, medical antecedents and surgery undergone. The MTS describes the evolution of physiological variables throughout the surgical operation. The EL records the medical actions performed on the patient during the surgery.

So far, no application-independent approach exists in the literature, that can achieve contextualized time series predictions informed by sequences of events. To fill this gap, we introduce the Non-Homogeneous Markov Chain AutoRegressive (NHMC-AR) model.

*Fatoumata Dama was supported by a PhD scholarship granted by the French Ministery for Higher Education, Research and Innovation. The experiments were performed at the CCIPL (Centre de Calcul Intensif des Pays de la Loire, Nantes, France).

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.



(a) NHMC-AR(q,$p = 2$) (b) PHMC-MLAR(q,$p = 2$)

Fig. 1: Conditional independence graphs for (a) the Non-Homogeneous Markov Chain AutoRegressive model and (b) the Partially Hidden Markov Chain Multivariate Linear AutoRegressive model. To simplify, we consider a univariate time series ($d = 1$). t: time; S_t: state; X_t: (observed) data measure point of the time series; C_t: (observed) contextual variable. Blue color: observed item; pink color: latent state. In (a), all states are latent. In (b), S_1 and S_t are latent. S_2 is observed. S_3 and S_{t-1} are partially observed (*i.e.*, the state value is known to belong to some subset of the set of all observable states, of size $\in [2,\, q-1]$).

2 Model definition

Let x be an MTS collected by d sensors. This MTS is a sequence of length T, $x_1^T = \{x_1, x_2, \ldots, x_T\}$, where $x_t \in \mathbb{R}^d$ is a d-dimension data point at time t ($t = 1, 2, \cdots T$). Let $\mathcal{E} = \{1, 2, \ldots, r\}$ be a finite set of r event categories, whose discrete events are likely to influence the behavior of the system considered. We can map each time $t \in [1,T]$ to a *time stamp* $\omega_t \in \mathbb{R}^+$, on the continuous time line. Let $\mathcal{L}_{og} = \{(e_i, t_i)\}_{e_i \in \mathcal{E}, i=1,\cdots N}$ ($t_i < t_{i+1} \leqslant \omega_T$) the log of time-stamped events observed in parallel to x_1^T. At a given time t, the history of past events of the MTS is the subsequence $\mathcal{H}_t$ of $\mathcal{L}_{og}$ whose events occurred at time stamps $t_i \leqslant \omega_t$.

The first innovation in NHMC-AR lies in the incorporation of contextual variables that allow the propagation of time-varying information from an EL to an observed time series x_1^T. Given the context $\mathcal{H}_t$, we must define a function Φ to calculate the contextual variable $c_t = \Phi(\mathcal{H}_t)$. We assume that Φ is constant over time. The second key innovation is the use of the Hawkes temporal point process framework [4] in the design of the contextual variables.

2.1 Three stochastic processes at work

Figure 1 (a) illustrates the two conditional independence assumptions that govern the relations within ($\{S_t\}$, $\{X_t\}$, $\{C_t\}$) in the NHMC-AR model. We denote by q the finite number of observable states of the system. The dependence of S_t on S_{t-1} and C_t characterizes a Markov chain (first-order hidden Markov model) which is non-homogeneous. The dependence of X_t on S_t and the p lagged variables X_{t-p}^{t-1} features a state-specific dynamics. The latter is described through a state-specific Linear AutoRegressive (LAR) model, in the NHMC-**AR** model.

2.2 Contextual variables

Recent works have highlighted the relevance of Hawkes temporal point processes (HTPPs) to extract features from sequences of events, for classification tasks [3].

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Let $\lambda(\omega \mid \mathcal{H}_t) = (\lambda_1(.), \lambda_2(.), \ldots, \lambda_r(.))$ the intensity function of the HTPP. Each term $\lambda_e(\omega)\delta t$ $(e = 1, 2 \cdots r)$ defines the probability that the event from category e occurs in interval $[\omega, \omega + \delta t[$, given the history of past events:

$$\lambda_e(\omega \mid \mathcal{H}_t) = \mu_e + \sum_{(e_i,t_i)\in\mathcal{H}_t} \zeta_{e,e_i}\, \rho \exp(-\rho(\omega - t_i)), \quad t_i \leqslant \omega_t \leqslant \omega,\ e, e_i \in \mathcal{E},$$

with $\mu_e, \zeta_{e,e_i}, \rho \in \mathbb{R}^+$, respectively the basic intensity for events of type e, the positive influence of e_i category's events on e category's, and a decay parameter.

We now make the link between Φ and the HTPP, and we define the contextual variable as $c_t = \Phi(\mathcal{H}_t) \in \mathbb{R}^r$, with $c_{t,e} = \lambda_e(\omega_t \mid \mathcal{H}_t), e = 1, 2, \cdots, r$.

2.3 Transition probabilities

We model the time-varying transition probability $b_{i,j}(t)$ from state i to state j as the interaction between two terms:

$$b_{i,j}(t) = P(S_t = j \mid S_{t-1} = i, C_t = c_t) = \frac{a_{i,j} \times \mathcal{G}(c_t; i, j)}{\sum_{k=1}^{q} a_{i,k} \times \mathcal{G}(c_t; i, k)} \quad t = 2, 3, \ldots$$

The term $a_{i,j} = P(S_t = j | S_{t-1} = i)$ denotes the standard transition probability in homogeneous Markov chains. The term $\mathcal{G}(c_t; i, j)$ quantifies the effect of the contextual variable c_t on transition $i \to j$. Since this term contributes to favor or impede transition $i \to j$, it also impacts the change in the behavior of the MTS. Function $\mathcal{G}$ is defined as a regression of the r components of the contextual variable:

$$\mathcal{G}(c_t; i, j) = \exp\left(\sum_{e=1}^{r} \eta_{i,j,e} \times c_{t,e}\right).$$

The parameter $\eta_{i,j,e} \in \mathbb{R}$ represents the specific impact of the event category e on the transition $i \to j$.

3 Parameter estimation

We denote by $\boldsymbol{\theta}^{(S)} = (\pi_i, a_{i,j}, \eta_{i,j,e})_{i,j=1,\ldots,q;e=1,\ldots,r}$ the parameters associated with the state process, where π_i is the probability that the initial state S_1 is i. The parameters associated with the emission probabilities $\big(P(X_t = x_t \mid X_{t-p}^{t-1}, S_t = k)\big)_{k=1,\ldots,q}$ are $\boldsymbol{\theta}^{(X)} = (\boldsymbol{\theta}^{(X,1)}, \ldots, \boldsymbol{\theta}^{(X,q)})$. The set $\boldsymbol{\theta}^{(X,k)}$ represents the parameters specific to the LAR model used for state k. To estimate $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(X)}, \boldsymbol{\theta}^{(S)})$ from a training set consisting of pairs of MTSs and ELs, we employ an Expectation-Maximization (EM) approach. EM iteratively computes estimates $\hat{\boldsymbol{\theta}}_n$ by alternating E and M steps until convergence:

$$\textbf{E step} \qquad Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{n-1}) = \mathbb{E}_{P(S_1^T \mid X_{1-p}^T, C_1^T; \hat{\boldsymbol{\theta}}_{n-1})}\left[\ln \mathcal{L}_c^c(\boldsymbol{\theta})\right]$$

$$\textbf{M step} \qquad \hat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{n-1}),$$

where X_{1-p}^0 represents the p initial values of the MTS $\{X_t\}$ and $P(S_1^T \mid X_{1-p}^T, C_1^T; \hat{\boldsymbol{\theta}}_{n-1})$ is the *a posteriori* law of the latent states S_1^T given the current approximation $\hat{\boldsymbol{\theta}}_{n-1}$. The term $\mathcal{L}_c^c(\boldsymbol{\theta}) = P(X_1^T = x_1^T, S_1^T = s_1^T \mid X_{1-p}^0, C_1^T = c_1^T; \boldsymbol{\theta})$ is the conditional likelihood of the completed data of the NHMC-AR model. The term

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

$\ln \mathcal{L}_c^c(\boldsymbol{\theta})$ is a random variable; the term Q represents its expectation with respect to the *a posteriori* law of S_1^T. We use the conditional independence graph of the model (Fig. 1 (a)) to decompose $\mathcal{L}_c^c$: $\mathcal{L}_c^c(\boldsymbol{\theta}) = \prod_{t=1}^{T} f_{s_t}(x_t)\ \pi_{s_1} \prod_{t=2}^{T} b_{s_{t-1},s_t}(t)$, where $f_k(x_t)$ is the emission probability in state k.

To perform the E-step, we adapted the standard forward-backward recursive algorithm used for the homogeneous Markov chains. Then, in a nutshell, once Q is calculated, its maximization divides into $q+1$ sub-problems whose object is to maximize $(Q_X^{(k)})_{k=1,2,\cdots,q}$ and Q_S. Under some precautions not detailed here, analytical expressions may be derived for the $Q_X^{(k)}$ terms. Maximizing Q_S requires numerical optimization methods (*e.g.*, quasi-Newton methods). The source codes for the NHMC-AR model are available at https://github.com/reviewerchris/repository_nhmc-ar_c1_c2_23_04_24_mon_20h40.

4 Evaluation of the model

4.1 Experimental settings

Dataset. We obtained access to data describing laparoscopic inguinal hernia surgery, for 1000 male patients around 30 years, and with no medical antecedents. The MTSs describe 4 physiological variables: heart frequency (HF); systolic, average and diastolic blood pressures (SBP, ABP, DBP). ABP is not computed as the mean of SBP and DBP but is measured separately. We consider a training, validation, and test set of sizes 500, 200 and 300 anaesthesia profiles, respectively.

We identified $r = 5$ categories of events: H (administration of hypnotics), M (administration of morphinics), 3 levels of painful *stimuli* (P_L: low, P_M: medium, P_H: high). The 5 categories are present in all the ELs of the data.

Extraction of the contextual variables defined *via* a Hawkes process. We employed the least squares method with Ridge regression to estimate the parameters $(\mu_e, \zeta_{e,e_i})_{e,e_i=1,2,\cdots,r}$ of the Hawkes process, using the training set. We varied the decay hyperparameter ρ in a grid search as well as the regularization parameter C. We selected the HTPP that yielded the highest prediction accuracy for the categories of events, on the validation set. Then we computed the contextual variables for the anaesthesia profiles of the training set.

Competing methods. We identified no application-independent model able to predict MTSs informed by ELs. Thus, we developed an alternative version of NHMC-AR, denoted **NHMC-AR-B**, in which the context variables are defined as $c_t = \Phi_B(\mathcal{H}_t) \in \mathbb{R}^{r+1}$, with $c_{t,r+1} = \omega_t$ and $c_{t,e} = max\ \{t_i \mid (e_i,\ t_i) \in \mathcal{H}_t \text{ and } e_i = e\}, e \in \mathcal{E}$. The term $\mathcal{G}_B(c_t; i, j)$ is an exponential function of the sum of weighted impacts of events and pairs of events (not shown). In contrast to NHMC-AR, the extraction of the contextual variables is straightforward.

The recent state-of-the-art model **PHMC-MLAR** (Partially Hidden Markov Chain Multivariate Linear AutoRegressive model) [2] (https://github.com/damaf/PHMC-MLAR) is an MSAR which is able to handle partially annotated states (see Figure 1 (b)). In our experiments, we use the partial state annotation principle to synchronize an MTS and its EL before training or using the model: at time

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Model	Number of categories of events (n_c)	Number of latent states (n_ℓ)	Total number of states (q)
PHMC-MLAR-5	5 {H, M, P_L, P_M, P_H}	0 to 1	5 to 6
PHMC-MLAR-4	4 {H, M, P_{LM}, P_H}	0 to 2	4 to 6
PHMC-MLAR-3	3 {H, M, P_{LMI}}	0 to 3	3 to 6

Table 1: Categories of events in the 3 PHMC-MLAR versions. ($q = n_c + n_\ell$)

t, if no event occurred in $]\omega_{t-1}, \omega_t]$, S_t is a latent state; otherwise, S_t is set as the subset of the categories of events that occurred in this interval. We chose to examine three versions of PHMC-MLAR: in PHMC-MLAR-5, we consider the 5 categories of events H, M, P_L, P_M and P_H. Table 1 shows how we merged painful *stimuli* to obtain the versions PHMC-MLAR-3 and PHMC-MLAR-4.

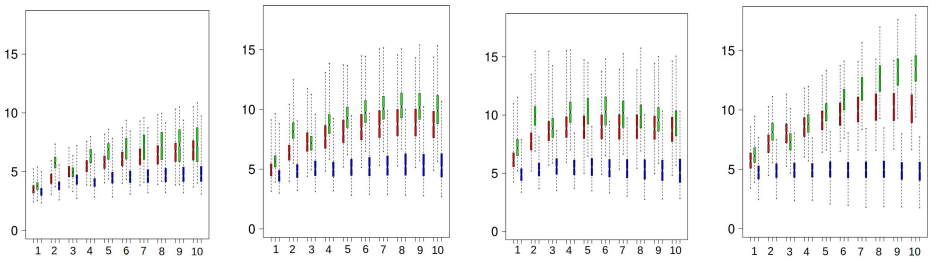
Evaluation metric. We assessed the predictive performance of trained models through the Mean Absolute Percentage Error (MAPE) defined as MAPE $= \frac{100}{N}\sum_{i=1}^{N}\left|\frac{o_i - p_i}{o_i}\right|$, where o_i and p_i are the observed and predicted values for ith prediction. Given a prediction horizon h and a physiological variable, we compute the MAPE value of a time series of length T by varying the prediction origin in $[p+1, T-h]$. We can compute an average MAPE on a validation or test set. By summing the average MAPEs of the 4 variables, we obtain a global predictive performance, further denoted MAPE-G. We made the predictions under the same condition as in the application targeted: no event will occur within the prediction horizon. To simulate a patient, we chain h-ahead predictions, at times t, $t+h$, $t+2h\cdots$. When a medical action is triggered, a h-ahead prediction is performed earlier than expected, which starts a new series of predictions.

4.2 Results

We varied the number of states q in [2,5] to select the 2 NHMC-AR models. We varied q as shown in Table 1 (3rd and 4th columns) for the 3 PHMC-MLAR versions. For all 5 models, the autoregressive order p was chosen in [0,5]. We first pre-selected the 6 configurations (q, p) showing the lowest BIC scores on the training set, for each of the 5 models. Then, we selected one configuration for each of NHMC-AR and NHMC-AR-B, and one from the 3 PHMC-MLAR versions: to this end, we computed the MAPE-G criterion at prediction horizon 10, on the validation set. The 3 selected models are PHMC-MLAR-4($q = 6$, $p = 2$), NHMC-AR-B($q = 4$, $p = 3$) and NHMC-AR($q = 4$, $p = 5$).

Figure 2 allows to compare the distributions of MAPE values across the 4 variables and the 10 horizons. For the PHMC-MLAR-4 selected, regardless of the prediction horizon, the average MAPE is lower than 6.8%, 9.0%, 9.0% and 10.3% for HF, SBP, ABP and DBP, respectively. These maxima are 7.3%, 10.4%, 10.4% and 13.5% for the NHMC-AR-B selected. The maxima obtained for the NHMC-AR selected are much lower: 4.9%, 5.4%, 5.6% and 5.1%. NHMC-AR is also the only model to present stable predictive performances over all horizons. Table 2 recapitulates the global predictive performances of the 3 models.

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.



(a) HF (b) SBP (c) ABP (d) DBP

Fig. 2: Predictive performances of the 3 selected models, for the 4 physiological variables, and for 10 prediction horizons. HF: heart frequency; SBP, ABP, DBP: systolic, average and diastolic blood pressures. Red: PHMC-MLAR. Green: NHMC-AR-B. Blue: NHMC-AR. MAPE values are computed over the test set. NHMC-AR shows the lowest MAPEs, which are stable across the horizons. Globally, PHMC-MLAR shows better performances than NHMC-AR-B.

Model	Horizon									
	1	2	3	4	5	6	7	8	9	10
PHMC-MLAR	20.6	25.8	29.0	30.8	32.1	33.3	34.1	34.7	34.7	34.6
NHMC-AR-B	23.0	32.1	28.4	34.3	36.3	38.1	39.4	40.3	40.5	40.2
NHMC-AR	17.5	19.0	20.1	19.7	20.5	20.5	20.7	20.8	20.7	20.6

Table 2: Global predictive performance (MAPE-G) of the 3 selected models, over the 4 physiological variables. MAPE-G is computed on the test set.

5 Conclusion

The NHMC-AR model is an extension of the hidden Markov Model in which the latent state process is conditioned on contextual variables. The latter are derived from the history of past events collected in parallel to the time series. The Hawkes point process framework is used for this purpose. NHMC-AR substantially outperforms a recent state-of-the art model adapted for event-based time series prediction, as well as another NHMC-AR instantiation. Future work will involve investigating other temporal processes than the Hawkes process.

References

[1] M. Alsefri, M. Sudell, M. García-Fiñana, and R. Kolamunnage-Dona, Bayesian joint modelling of longitudinal and time to event data: a methodological review, *BMC Medical Research Methodology*, 20(1):1–17, 2020.

[2] F. Dama and C. Sinoquet, Partially Hidden Markov Chain Multivariate Linear Autoregressive model: inference and forecasting - application to machine health prognostics, *Machine Learning*, 112(1):45–97, 2023.

[3] H.S. Dutta, V.R. Dutta, A. Adhikary and T. Chakraborty, HawkesEye: detecting fake retweeters using Hawkes process and topic modeling, *IEEE Transactions on Information Forensics and Security*, 15:2667–2678, 2020.

[4] A.G. Hawkes, Spectra of some self-exciting and mutually exciting point processes, *Biometrika*, 58:83–90, 1971.