Evaluation of Minimal Data Size by Using Entropy, in a HMM Maintenance Manufacturing Use

Bernard Roblès, Manuel Avila, Florent Duculty, Pascal Vrignat, Stéphane Begot, Frédéric Kratz

Evaluation of Minimal Data Size by Using Entropy, in a HMM Maintenance Manufacturing Use

Bernard Roblès * **Manuel Avila** * **Florent Duculty** *
Pascal Vrignat * **Stéphane Bégot** * **Frédéric Kratz** **

* *University of Orleans, PRISME Laboratory IRAuS team, 2 Avenue Mitterrand 36000 Châteauroux, FRANCE (e-mail: Bernard.Robles@univ-orleans.fr, Manuel.Avila@univ-orleans.fr, Florent.Duculty@univ-orleans.fr, Pascal.Vrignat@univ-orleans.fr, Stephane.Begot@univ-orleans.fr).*
** *PRISME Laboratory IRAuS team, ENSI 88 boulevard Lahitolle 18020 Bourges Cedex, FRANCE (e-mail: Frederic.Kratz@ensi-bourges.fr).*

Abstract: In this paper, we wish to find a minimal data size in order to better conceptualize industrial maintenance activities. We based our study on data given by a Synthetic Hidden Markov Model. This synthetic model is intended to produce real industrial maintenance observations (or "symbols"), with a corresponding degradation indicator. These time series events are shown as Markov chains, also called "signatures". The production of symbols is generated by using a uniform and a normal distribution. The evaluation is made by applying Shannon entropy on the HMM parameters. The results show a minimal number of data for each distribution studied. After a discussion about the use of a new "Sliding Window" of symbols usable in a Computerized Maintenance Management System, we developed two industrial applications and compare them with the best optimized "signature" previously found.

Keywords: Hidden Markov Models, Shannon entropy, model optimization, learning algorithm, predictive maintenance.

1. INTRODUCTION

Hidden Markov Models (HMM) can be used for modeling complex systems. In this study, we use this kind of model to detect faults in systems. Nowadays, industrial robots operating in a stochastic environment need upstream fault detection in order to prevent any breakdowns. Indeed, it is conceivable for poorly maintained equipment to break down, bringing the entire production line to a halt. In order to help maintenance experts, we built a *Synthetic Model* based on a *reference model*, which fits real industrial processes (Vrignat et al. (2010)). We showed in Vrignat et al. (2010), that this *reference model* provided good failure prediction. This *Synthetic Model* produced observations or symbols, commonly used in a CMMS [1] . It also produced degradation indicators: S1, S2, S3 and S4 depicted in Figure 1. We have previously found in Roblès et al. (2012) the best HMM topology usable in industrial maintenance (see Figure 2).
Most published papers aim to optimize stochastic problems by using the principle of minimum Cross-Entropy (Shahid (2012) and Botev et al. (2007)). Indeed, this method introduced first by Rubinstein (1997) is usually applied to combinatorial optimization problems (e.g. the traveling salesman problem).
In this paper, we attempt to find the minimal number of data necessary to have the best relevancy of the HMM parameters. In order to find this minimal data size, we try to find a maximum entropy value or a convergence entropy limit on from the synthetic industrial databases. Different kinds of distributions are tested: uniform and normal distributions. Assessment of the produced observations was made by using criteria usually used in model selection: the Shannon Entropy.
The structure of this paper is as follows: in Section 2 we outline HMM, present methods to assess the relevancy of model parameters and describe the Synthetic Model design. The results are given in Section 3 and discussed in Section 4. Finally the findings are compared with two industrial applications, in Section 4.3.
The purpose of this study is to find a minimal data size given by a Synthetic Hidden Markov Model. We would like to show that our minimal size choice meets objectives in the improvement of preventive maintenance and breakdown prediction, in the industrial sector.

2. METHODOLOGY

2.1 Hidden Markov Model

In this study, we chose HMM to describe industrial maintenance events. An HMM consists of a hidden stochastic

[1] A Computerized Maintenance Management System is an information database about maintenance operations. This information is intended to make a decision support for maintenance experts.

process modeled by a Markov chain and an observable stochastic process. This kind of model is represented by automaton with hidden states which consists of unobservable variables (Rabiner (1989), Fox et al. (2006)). These unobservable variables represent the system status to be modeled. Only output variables are observable. Moreover, this automaton is intended to generate observation sequences from its outputs (see an example of model topology depicted in Figure 1). Indeed, we attempted to assess the minimal data number for a HMM use.

HMM is characterized by:

- State number;
- Number of distinct observation symbols per state, observation symbols corresponding to the physical output of the system being modeled;
- Distribution probability of state transitions;
- Distribution probability of observation symbols;
- Initial states distribution.

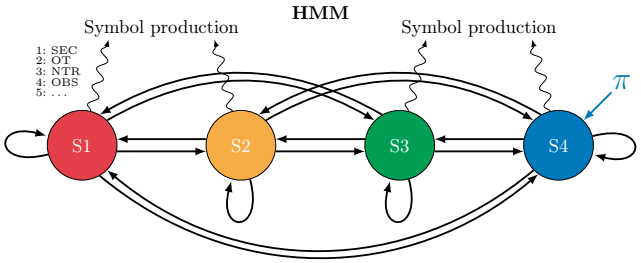

Fig. 1. A four-state discrete HMM. A discrete HMM consists of a hidden stochastic process modeled by a Markov chain and an observable stochastic process whose states dependent on the hidden process. This model is a stochastic automaton. It represents the degradation level (S1 to S4) of an industrial process. S1 state: the system is stopped. S4 to S2 state: progressive degradations of the process. This is the general case, when all transitions are possible. This kind of automaton produces observation sequences or symbols. These symbols represent the Markov chain given by a maintenance database.

Markov assumption States prediction is not made any more accurate by adding a priori knowledge of the information, i.e. all useful information for future prediction is contained in the present state of the process (*i.e.* it's a Markov chain of order 1).

$$P(X_{n+1} = j | X_0, X_1, \ldots, X_n = i) = P(X_{n+1} = j | X_n = i). \quad (1)$$

We used this assumption to take into account several orders of a Markov chain.

Definitions of a discrete HMM We describe here the variables used in an HMM:

- Let N, be the number of workable hidden states and $\mathbb{S} = \{s_1, s_2, \ldots, s_N\}$, the set of states. Let q_t, be the value of this variable at time t;
- The modeled process must match first-order Markov assumption (see Eq. 1);
- Let T, be the full number of observation symbols and let $X = \{x_1, x_2, \ldots, x_T\}$, observations sequence of the modeled process;
- Let $A = \{a_{ij}\}$, distribution probability of state transitions where:

$$a_{ij} = P(q_{t+1} = s_j | q_t = s_i) \quad 1 \leqslant i, j \leqslant N, \quad (2)$$

- Let $B = \{b_j(m)\}$ distribution probability of observation symbols in j state, where:

$$b_j(m) = P(X_t = x_m | q_t = s_j) \quad 1 \leqslant j \leqslant N \;\; 1 \leqslant m \leqslant T, \quad (3)$$

  where X_t, is the value of observation variable at time t.
- Let $\pi = \{\pi_i\}$, initial states distribution where:

$$\pi = P(q_1 = s_i) \quad 1 \leqslant i \leqslant N, \quad (4)$$

- HMM will be set as: (A, B, π).

2.2 Shannon entropy

This criterion was used in Roblès et al. (2012), to evaluate relevancy of the "signatures" and find the best model studied. In this paper, we calculated the Shannon entropy in the 1^{st} order Markov chains for simulated data and real industrial data.
Shannon entropy is a function which calculates the information rate contained in an information source. This source can be a text written in any language, an electrical signal or an unspecified electronic file.

Entropy definition Shannon entropy is defined in Cover and Thomas (1991) as follows:

$$H(S) = -\sum_{i=1}^{n} P_i \log_{10} P_i, \quad (5)$$

P_i is the average probability to find the i symbol in S.

The two principles of entropy maximization in Agouzal and Lafouge (2008) are the following:

- The principle of probabilities assignment to a distribution when we haven't enough information on it;
- For all probability distributions that satisfy the constraints, we choose the one which has the maximum entropy according to Shannon.

The 2^{nd} principle is used in Chandrasekaran et al. (2007) for model selection, and Arminjon and Imbault (2000) for building even more accurate models, by adding information. Our approach consists in calculating the entropy evolution in order to find a maximum entropy or an asymptotic entropy value. The minimal number of data will then appear for different distributions and for the industrial cases.
Beyond model selection, we used these two principles to eliminate outlier symbols. Indeed, some observations might be due to human input mistake. We therefore eliminated these symbols an try to find a convergence entropy with the others.

The convergence entropy value (L) is defined by the limit of entropy function of S, as S (the observations serie) approaches $+\infty$:

$$\lim_{S \to +\infty} H(S) = L \tag{6}$$

2.3 Evaluation process

We have used synthetic model to produce about 1000 data events. These simulated symbols, according to real industrial process, are obtained by using uniform and normal distribution. Correlatively, we produced states for others topologies by using the same process. Afterwards, these states are used to compare states models. Insofar as states are obtained by different learning and decoding algorithms:

- Baum-Welch learning (Baum et al. (1970)), decoding by Forward Variable (Rabiner (1989)),
- Segmental K-means learning (Juang and Rabiner (1990)), decoding by Viterbi, (Viterbi (1967)).

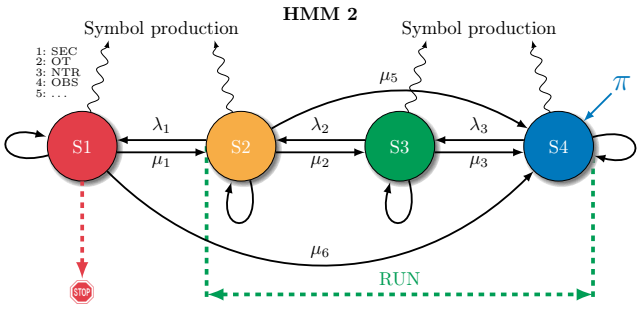

Fig. 2. The four-state Hidden Markov Model. S1 to S4 represent the degradation levels. S1: the system is stopped. S4 to S2: progressive degradations of the process. The production of symbols represents the Markov chain given by a maintenance database. λ_i are failure rates and μ_i are repair rate. π is the initialization matrix.

2.4 HMM topology

We used the HMM topology (see Figure 2), represented by automata with four oriented states. These stochastic automata, depicted in Figure 2, represent the degradation level of an industrial process, S4 to S1. {S4, S3, S2} states, when the process is running ("RUN"), and {S1} state, when the process is stopped ("STOP"). With this topology we need to go through all states (S2 and S3) to go from a high level of availability (S4) to a low level of availability (S1). S1 is the state of a breakdown.
Regular temporal sampling is a requisite to have a Markov process. Both the simulated process and industrial database must have the same temporal sampling. We can therefore use Markov modeling.

2.5 Simulated industrial CMMS

Nowadays, many industrial factories use **preventive maintenance**. Maintenance operators consign their actions and observations in a centralized database. We show an example of such database in Table 1. For instance, symbols "PM, OT, SP, ..." could characterize maintenance activities carried out on industrial processes. We recall the meaning of selected symbols resulting from observations

Name	Date	Ope.	Cd	IT	N°	Code
Dupond	11/01/2007	Lubrication	PM	20	1	9
Dupond	11/01/2007	Lubrication	PM	20	2	9
Dupond	12/01/2007	Lubrication	SEC	30	3	5
Dupond	12/01/2007	Lubrication	PM	30	4	5
Dupond	13/01/2007	Padlock	PM	10	5	6
Dupond	13/01/2007	Padlock	NTR	30	6	5
Dupond	16/01/2007	Lubrication	SP	90	8	1
Dupond	19/01/2007	Padlock	OT	10	9	3
...	...	...	...	...	...	...

Table 1. Example of recorded events from a maintenance database.

given in Table 2. The "SP" symbol corresponds to a stop of production units: process state = "STOP" in Table 2. It is a critical condition, which our research aims to minimize. Process state = "RUN" when production units are running without failure.
We conceptualized this kind of maintenance here by using

N°Obs.	Symbols	Interventions type
1	SP	(Troubleshooting / Stop Production)
2	SM	(Setting Machine)
3	OT	(Other)
4	OBS	(Observation)
5	PM	(Preventive Maintenance)
6	SEC	(Security)
7	PUP	(Planified Upgrading)
8	CM	(Cleaning Machine)
9	PMV	(Preventive Maintenance Visit)
10	NTR	(Nothing to report)

Process states
STOP
RUN

Table 2. Symbolic coding system of maintenance interventions.

a **Synthetic Model** presented in the next paragraph (§2.6) to simulate this real industrial environment. We chose "λ_i" (failure rate) and "μ_i" (repair rate) of HMM parameters (Vrignat et al. (2010)), to match the maintenance register as closely as possible (Table 1).

2.6 Conceptual Synthetic Model

We designed a Synthetic Model with Matlab by using four-state oriented HMM presented in Figure 2. This choice is discussed in the paragraph § 4.
HMM topology only depends on matrix elements, where $\{a_{ij}\} \neq 0$ (If all matrix elements are different from zero, we have a "total connectivity matrix"). The transition matrix (A) has been specified in Vrignat et al. (2010) by:

$$A = \begin{pmatrix} 0.500 & 0.250 & 0 & 0.250 \\ 0.100 & 0.070 & 0.500 & 0.330 \\ 0 & 0.005 & 0.495 & 0.500 \\ 0 & 0 & 0.001 & 0.999 \end{pmatrix}. \tag{7}$$

The Synthetic Model built sequences of data (also named "signatures") by using 2 discrete distributions:

- Uniform distribution: This law makes it possible to model random variables uniformly distributed, as shown in Figure 5. A random variable of n values $\{x_1, x_2, \ldots, x_n\}$ has equal probability of:

$$P(x_i)_{i \in \{1,\ldots,n\}} = \frac{1}{n}. \tag{8}$$

- Normal distribution (Laplace – Gauss): This law makes it possible to model many distributions of random variables. This distribution is defined by the probability density function $\varphi : \mathbb{R} \rightarrow \mathbb{R}^+$, where parameter μ is expectation, and σ the standard deviation (See Figure 4):

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} \, \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \quad (9)$$

We used these symbol sequences as a Markov chain (see Table 3), to model the degradation level of a process.

PM	PM	SEC	PM	PM	NTR	NTR	SP	...

Table 3. Sequence of a message from maintenance database.

We will now describe the entropy assessment specifications used for data analysis (see Figure 3). We produced 12 sequences distributed among:

- 1000 2-tuple $(Symb\_U, State\_U)$, for a Uniform distribution (Figure 5),
- 1000 2-tuple $(Symb\_N, State\_N)$, for a Normal law (Figure 4).

Each sequence ends with a production stop (symbol SP in red).

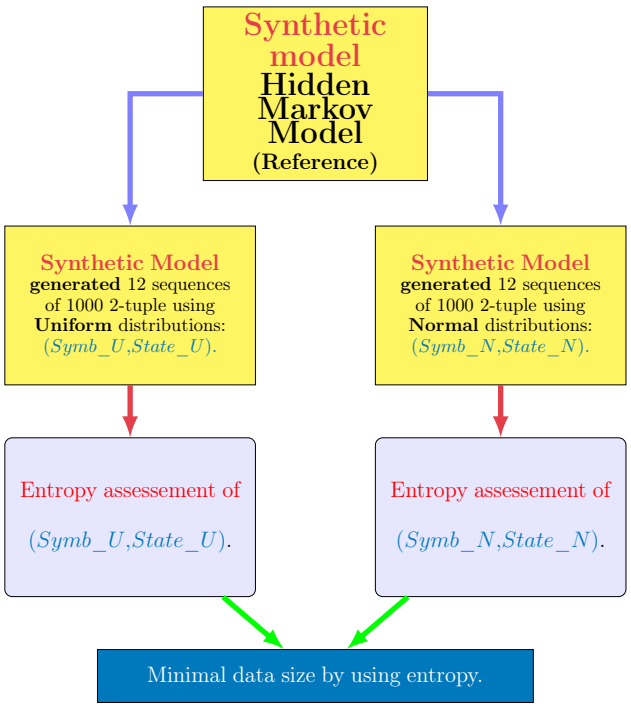

Fig. 3. Entropy assessment specifications: The Synthetic Model generated 1000 $(Symbols, States)$ by using the normal and the uniform distribution $((Symb\_U, State\_U)$ and $(Symb\_N, State\_N))$. All 2-tuple are distributed among the 12 sequences. The number of the estimated 2-tuple are assessed to determine the minimal size of data by using Shannon Entropy.

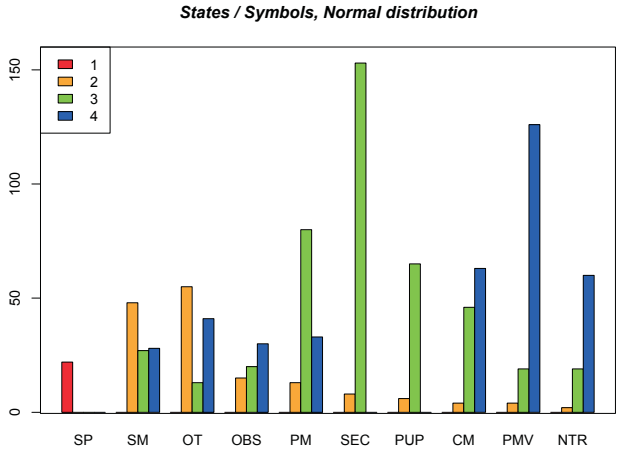

Fig. 4. The Normal distribution of symbols, given by Hidden Markov Model reference. Stop Production symbol (SP) represents the state 1 of the model. (OT) symbol (OTher) is the most representative of state 2. (SECurity) is the most representative of state 3, Preventive Maintenance Visit (PMV) is the most representative of state 4.

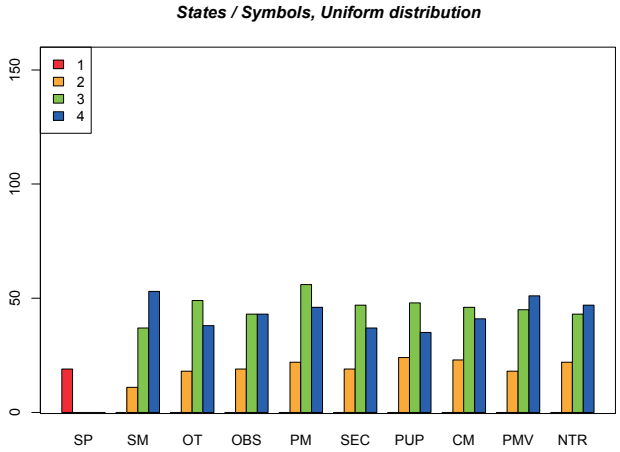

Fig. 5. The Uniform distribution of symbols, given by Hidden Markov Model reference.

2.7 Industrial applications

This approach involves optimizing the **signatures** of any industrial CMMS by using HMM topologies. Furthermore, our minimal data size would be able to provide best decision support for organizing daily maintenance and would help experts to improve maintenance activities.

Context: The industrial cases consist in studying a continuous process of bread production (HARRYS) and a glass factory (Arc International). When any failure occurs on the subsystems of the line, that involves stopping the entire production line (a lot of bakery products are wasted). Therefore, preventive maintenance has been scheduled to prevent such cases. The factory operates all over the year, without any interruption. Teams are orga-

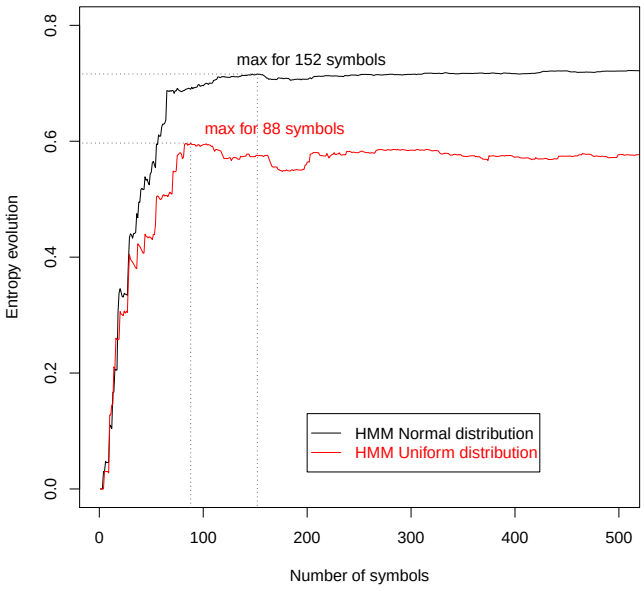

Fig. 6. Shannon Entropy evolution according to the number of data. We used the Synthetic Model to simulate data. These simulated symbols are obtained by using uniform and normal distribution. Entropy is maximum for 152 symbols with the Normal distribution and for 88 symbols with the Uniform distribution.

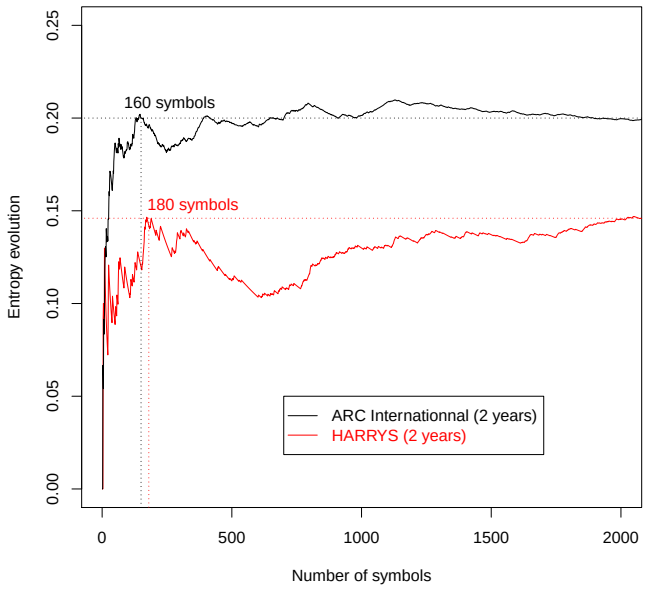

Fig. 7. Shannon Entropy evolution according to the number of data from the industrial cases: HARRYS and ARC International. The industrial data was collected from the years 2005 to 2007. Entropy obtained the asymptotic result for 180 symbols with the HARRYS database. The asymptotic value reach its maximum with 160 symbols for the ARC International database.

nized in shifts. Every maintenance operation or problem in the process must be recorded in the CMMS database. According to the internal maintenance policy, one day sampling is chosen. This regular temporal sampling makes it possible to be in a Markov process situation. We made comparisons with data for a period of 2 years. We had more than 2000 records from their database (see Table 1).

3. RESULTS

3.1 Synthetic model

The Synthetic Model is used to simulate data and produce symbols by using the Normal and the Uniform distribution. Shannon entropy evolution, depicted in Figure 6, shows a maximum for **152 symbols** with the Normal distribution and for **88 symbols** with the Uniform distribution. We also outlined an entropy value limit for each distributions, which shows that the number of symbols necessary to describe the maintenance activities is finite.

3.2 Entropy in industrial applications

Data comes from the HARRYS and the ARC International CMMS database. Data was collected from the years 2005 to 2007. Shannon entropy evolution according to the number of data is depicted in Figure 7. This figure shows an asymptotic value for both industrial cases. The entropy reached the asymptotic value with 180 symbols for the HARRYS database and with 160 symbols for the ARC International database.

4. DISCUSSION

The purpose of the study was to find a minimal data size given by a Synthetic Hidden Markov Model and compare findings with the two industrial applications.

4.1 Maximum entropy

Without a priori knowledge, we evaluated the relevancy of the signatures by measuring the Shannon entropy. We considered this signature as a 1^{st} order Markov Chain. As we showed in Roblès et al. (2011), Shannon entropy is a good indicator for model relevancy.
We found a maximum entropy value and results Figure 6 show an asymptotic value for the two distributions. The first peak of the entropy value for the uniform distribution is reached with 88 symbols and with 152 symbols for the normal distribution. These values are also the asymptotic entropy limit. It means that a finite number of symbols is necessary to describe the maintenance activities.
For predictive maintenance use, we could establish a "sliding window" containing the number of symbols previously found.

4.2 Sliding Window

The sliding window (see Figure 8) contains the minimal number of symbols found in Figures 6 and 7, where entropy is maximal. All symbols inside the sliding windows would be able to provide decision support for organizing

daily maintenance and would help experts to schedule maintenance activities. We found the minimal size of data that should be implemented in the learning algorithms, in order to re-evaluate the model. Indeed, at the sample time "Now", we do not know the next observation. With this sliding window, we would be able to estimate the next hidden states. The main purpose of the sliding window is to provide an update of the models occasionally.

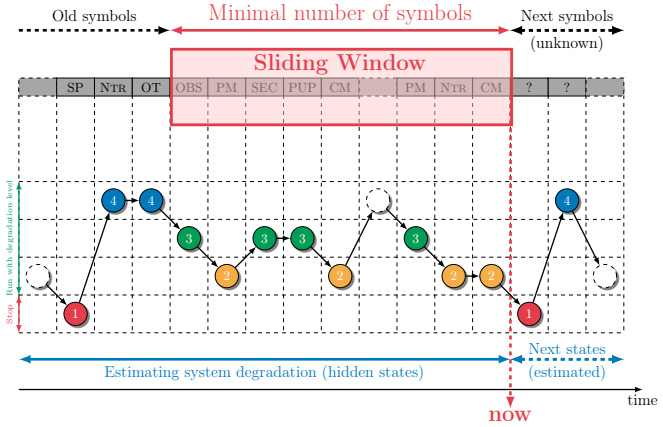

Fig. 8. The sliding window contains the minimal number of symbols, where entropy is maximal. All symbols inside the sliding windows would be able to provide decision support for organizing daily maintenance and would help experts to schedule maintenance activities. The main purpose of the sliding window is to provide an update of the models.

4.3 Confrontation of the Synthetic Model with two industrial applications

To confront the Synthetic Model with the real industrial cases, we compared data from industrial CMMS, with data given by the Synthetic Model: *Symb_U* and *Symb_N* (see Figure 3). The two tested distributions are the normal and the uniform ones.

In the two industrial cases studied (see the industrial contexts in Section 2.7.1), 160 symbols for ARC international and 180 symbols for HARRYS are necessary to find the maximum entropy value. As we see in Figure 6 and 7, the "industrial" entropy has almost the same evolution than the synthetic model. Moreover, the first maximum entropy value corresponding to the asymptotic entropy value. With these two numbers found, we could establish a minimal sliding window in order to optimize the maintenance activities in industrial sector. We showed that we can find a minimal number of data to properly estimate a model. Indeed these sliding windows from industrial database should determine the best number of symbols to conceptualize a predictive maintenance model.

4.4 Conclusion

The approach outlined in this study should be replicated with other subject areas with other models, as well as at other entropy orders. In an operational phase of industrial production, we will then re-evaluate the model regularly insofar the scheduled actions will change the behavior of the system.

REFERENCES

Agouzal, A. and Lafouge, T. (2008). On the relation between the maximum entropy principle and the principle of least effort: The continuous case. *J. Informetrics*, 2(1), 75–88.

Arminjon, M. and Imbault, D. (2000). Maximum entropy principle and texture formation. *Zeitschrift für angewandte Mathematik und Mechanik, 80, Suppl. N°1*, 13–16.

Baum, L.E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1), 164–171.

Botev, Z., Kroese, D., and Taimre, T. (2007). Generalized cross-entropy methods with applications to rare-event simulation and optimization. *Simulation*, 83(11), 785–806.

Chandrasekaran, V., Johnson, J.K., and Willsky, A.S. (2007). Maximum entropy relaxation for graphical model selection given inconsistent statistics. *Laboratory for Information and Decision Systems, Massachusetts Institute of Technology Cambridge, MA 02139.*

Cover, T.M. and Thomas, J.A. (1991). *Elements of information theory.* Wiley-Interscience, New York, NY, USA.

Fox, M., Ghallab, M., Infantes, G., and Long, D. (2006). Robot introspection through learned hidden Markov models. *Artif. Intell.*, 170(2), 59–113.

Juang, B.H. and Rabiner, L.R. (1990). The segmental k-means algorithm for estimating parameters of hidden Markov models. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 38(9).

Rabiner, L.R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceeding of the IEEE, 77(2) SIAM interdisciplinary journal*, 257–286.

Roblès, B., Avila, M., Duculty, F., Vrignat, P., and Kratz, F. (2011). Evaluation of relevance of stochastic parameters on Hidden Markov Models. In *Advances in Safety, Reliability and Risk Management, European Safety and Reliability Conference (ESREL), ISBN : 978-0-415-68379-1*, 71. Taylor & Francis Group, Troyes, France.

Roblès, B., Avila, M., Duculty, F., Vrignat, P., and Kratz, F. (2012). Statistical evaluation of Hidden Markov Models topologies, based on industrial synthetic model. In A.D.I.F.F.G. Borangiu Theodor; Dolgui (ed.), *14th IFAC Symposium on Information Control Problems in Manufacturing Information Control Problems in Manufacturing*, volume 14, 1029–1034. Bucharest, Roumanie.

Rubinstein, R.Y. (1997). Optimization of computer simulation models with rare events. *European Journal of Operations Research*, 99, 89–112.

Shahid, M.A. (2012). Cross entropy minimization for efficient estimation of sram failure rate. In W. Rosenstiel and L. Thiele (eds.), *DATE*, 230–235. IEEE.

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2), 260–269.

Vrignat, P., Avila, M., Duculty, F., and Kratz, F. (2010). Use of HMM for evaluation of maintenance activities. *IJAIS, International Journal of Adaptive and Innovative Systems, Vol. 1, Nos. 3/4*, 216–232.