Methods to choose the best Hidden Markov Model topology for improving maintenance policy

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

▶ **To cite this version:**

Bernard Robles, Manuel Avila, Florent Duculty, Pascal Vrignat, Stéphane Begot, et al.. Methods to choose the best Hidden Markov Model topology for improving maintenance policy. 9th International Conference on Modeling, Optimization & SIMulation, Jun 2012, Bordeaux, France. hal-00728654

HAL Id: hal-00728654
https://hal.science/hal-00728654

Submitted on 30 Aug 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

9^{th} *International Conference of Modeling, Optimization and Simulation - MOSIM'12*
June 6-8, 2012 - Bordeaux - France
"Performance, interoperability and safety for sustainable development"

Methods to choose the best Hidden Markov Model topology for improving maintenance policy

B. ROBLÈS, M. AVILA, F. DUCULTY, P. VRIGNAT, S. BÉGOT,

PRISME Laboratory
IUT de l'Indre
2 Avenue Mitterrand
36000 Châteauroux - France
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

F. KRATZ

PRISME Laboratory
ENSI 88 boulevard Lahitolle
18020 Bourges - France
Frederic.Kratz@ensi-bourges.fr

ABSTRACT: *Prediction of physical particular phenomenon is based on partial knowledge of this phenomenon. Theses knowledges help us to conceptualize this phenomenon according to different models. Hidden Markov Models (HMM) can be used for modeling complex processes. We use this kind of models as tool for fault diagnosis systems. Nowadays, industrial robots living in stochastic environment need faults detection to prevent any breakdown. In this paper, we wish to find the best Hidden Markov Model topologies to be used in predictive maintenance system. To this end, we use a synthetic Hidden Markov Model in order to simulate a real industrial CMMS**. *In a stochastic way, we evaluate relevance of Hidden Markov Models parameters, without a priori knowledges. After a brief presentation of a Hidden Markov Model, we present the most used selection criteria of models in current literature. We support our study by an example of simulated industrial process by using our synthetic model. Therefore, we evaluate output parameters of the various tested models on this process: topologies, learning algorithms, observations distributions, epistemic uncertainties. Finally, we come up with the best model which will be used to improve maintenance policy and worker safety.*

KEYWORDS: *Hidden Markov Models, model selection, learning algorithms, statistical test, uncertainties, predictive maintenance.*

1 INTRODUCTION

According to the Global Energy Statistical Yearbook 2011, *Enerdata*[1] gives alarming conclusions: after the 1% decrease observed in 2009, energy consumption soared by 5.5% in 2010, and results in a growth in CO_2 energy emissions close to 6%, to their highest level ever.

Despite heavy investment to remain competitive, most industry are not concerned with the *green thinking*, whereas implementing energy reducing measures such as having an efficient maintenance policy, is not so expensive and can save in energy costs. Obviously, fault diagnostics techniques can reduce maintenance downtime and thus reduce consumption of energy. According to (Vrignat, Avila, Duculty & Kratz 2010), we find two keywords in maintenance definition: *maintain* and *restore*. The first one refers to preventive action. The second refers to corrective action. Thus, maintenance optimization for reliability determines "optimal" preventive maintenance. Events preceding a problem in maintenance activities are often recurrent. Special events series should inform us on next failure. For example, in mechanical systems, noises, vibrations precede a failure. The loss of performances reflects failure or technical faults. We also show in (Vrignat et al. 2010) that our model provides a good failure prediction. We make a reference model, named *synthetic model*, which fits to real industrial processes. Our research consists in evaluating different Hidden Markov Models topologies, with parameters outcoming from this industrial *synthetic model*. In this work, the emphasis is on measuring relevance of Hidden Markov Models parameters, based on several criteria used in current literature. Then, we try to give the best HMM topology. The structure of this paper is as follows: in section 2, we outline

*Computerized Maintenance Management System

[1] An independent information and consulting company specializing in global energy

hidden Markov model and define its parameters. After determining the stochastic nature of our synthetic model, we present criteria used to evaluate relevance and uncertainties of HMM, in section 3. We show the evaluation process in section 4. Finally, we use our synthetic model to compare several HMM topologies, from among a candidate set with previous criterion and try to give the best one, in section 5.

2 HIDDEN MARKOV MODEL

Hidden Markov Model ((Rabiner 1989), (Fox, Ghallab, Infantes & Long 2006)) is an automaton with hidden states which consists of unobservable variable. This one represents the system status to be modeled. Only output variable is observable. Moreover, we get observations sequence from output of the automaton. From now, we rename observations sequence as *symbols*, representing these observations (see an example of model topology in figure 1). This is precisely relevance of these symbols that we attempt to evaluate. Hidden Markov Model 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.

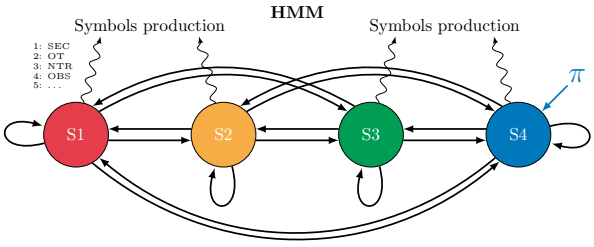


Figure 1: Four states Hidden Markov Model.

2.1 Markov Assumption

States prediction is not made more accurate by additional a priori knowledge information, i.e. all useful information for future prediction is contained in present state of the process.

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

2.2 Definitions for discrete Hidden Markov Model

Let us describe variables for HMM:

- Let N, the number of workable hidden states and $\mathbb{S} = \{s_1, s_2, \ldots, s_N\}$, the set of this variable. Let q_t, the value of this variable at time t;
- Modeled process, must match to first-order Markov assumption (see §2.1);
- Let T, 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 with:

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

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

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

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

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

- Hidden Markov Model will be set as: (A, B, π).

2.3 Learning algorithms and decoding algorithms

To achieve learning models, we use two different algorithms:

- Baum-Welch learning (Baum, Petrie, Soules & Weiss 1970), decoding by Forward Variable (Rabiner 1989).

  Estimate iteratively $\eta = (A, B, \pi)$, with an observation sequence of $X = \{x_1, x_2, \ldots, x_T\}$,

$$\text{Maximize} \rightarrow P(U = X | \eta), \quad (5)$$

- Segmental K-means learning (Juang & Rabiner 1990), decoding by Viterbi (Vrignat, Avila, Duculty, Aupetit, Slimane & Kratz 2011).

$$\text{Optimizing probability} \rightarrow P(X, S = Q^* | \eta). \quad (6)$$

  Q^*: Sequence of hidden states that most likely generated the sequence as calculated by the Viterbi algorithm. $S = (S_1, \ldots, S_T)$ is a T tuple of random values defined on $\mathbb{S}$

- Decoding algorithm by Forward Variable:

$$\alpha_t(j) = P(x_1, x_2, \ldots, x_t, Q_t = s_j|\eta). \quad (7)$$

- Viterbi decoding algorithm:

$$\delta_t(j) = max_{(q_1,\ldots,q_{t-1}\in\mathbb{S}^{t-1})}\{P(S_1 = q_1, \ldots, S_{t-1} = q_{t-1}, S_t = s_j), U_1 = x_1, \ldots, U_t = x_t|\eta\}. \quad (8)$$

3 EVALUATION METHODS

A lot of criteria in model selection are proposed in literature. We try to evaluate the best Hidden Markov Model topology proposed in (Vrignat et al. 2010), by using *Shannon's entropy* (Hocker, Xiaohu & Iyengar 2011), especially *maximum entropy principle* used in (Chandrasekaran, Johnson & Willsky 2007). Calculation is made with states and observations: symbols productions of synthetic HMM. To emphasize our analysis, we also use some criteria which penalize likelihood value, in order to overcome over-parameterization models, like *Akaike* (AIC) (Shang & Cavanaugh 2008) and *Bayes* (BIC) (Chen & Gopalakrishnan 1998) criteria. We begin to determine the stochastic nature of our given symbols.

3.1 NIST[2] Tests

First of all, we have to establish that we use stochastic density of probability. In (Rukhin, Soto, Nechvatal, Barker, Leigh, Levenson, Banks, Heckert, Dray, Vo, Rukhin, Soto, Smid, Leigh, Vangel, Heckert, Dray & Iii 2010), the authors propose a statistical package of 15 different tests. These tests were developed to test randomness of random number generators. NIST has verified the performance of these tests using a Kolmogorov-Smirnov test of uniformity on the p-values[3] (see § 3.6.1). The purpose of this test is to determine whether the number of ones and zeros in a sequence is approximately the same as would be expected to a truly random sequence. In our study, we use the *frequency test* of the NIST. This test validates that our synthetic model gives real stochastic symbols. The Decision Rule of the test, at the 1% Level, is: if the computed p-values is < 0.01, then we conclude that the sequence is non-random. Otherwise, we conclude that the sequence is random.

3.2 Shannon's entropy

We now study notions of Shannon's entropy. It is a function which calculate 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...

3.2.1 Entropy Definition

Shannon's entropy is defined in (Cover & Thomas 1991) as follows:

$$H(S) = -\sum_{i=1}^{n} P_i \log_b P_i, \quad (9)$$

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

3.2.2 Maximum entropy principles

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

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

(Chandrasekaran et al. 2007) use this 2nd principle for models selection, and (Arminjon & Imbault 2000) for building even more accurate models, by adding information. Our step consists in comparing the average entropies for various models. Value of average entropy would be then maximum for the most relevant model.

3.2.3 Entropic Filter

We now introduce "Entropic Filter" concept. According to the 2nd principle of entropy stated in §3.2.2, we choose the model whose average entropy is maximum. On the other side, outliers values can generate miscalculation in real entropy value of the model. Especially NTR symbols (Nothing To Report) which are not useful for evaluation (entropy is maximum). SP (Stop Production) symbols have likewise been eliminated (entropy is null). Indeed, they are totally discriminated for $S1$ state of HMM. To improve calculation of entropy, it is therefore better to eliminate these values. This approach is used through ID3 and C4.5 (Quinlan 1993) algorithm when creating decision tree, removing recursively attribute with zero entropy. In order to improve the calculation of entropy, we propose to eliminate discriminated symbols of zero entropy and the most representative symbols, where entropy is maximum. This operation will be

[2]National Institute of Standards and Technology

[3]The probability (under the null hypothesis of randomness) that the chosen test statistic will assume values that are equal to or worse than the observed test statistic value when considering the null hypothesis. The p-value is frequently called the "tail probability".

named "Entropic Filter". We then calculate the average entropy of models to assess relevance of observation sequences. The best model is the one which has the best average entropy, after entropic filtering.

3.3 Maximum likelihood

Let us now turn to studying maximum likelihood principle. Let P_α, a statistical model, and X, an observation sequence, the probability to see X according to P can be measured by $f(X, \alpha)$ function which represents the density of X when α appears. Since α is unknown, it seems natural to promote values of α where $f(X, \alpha)$ is maximum: it is the notion of likelihood of α for observation X.

– Expression of likelihood V:

$$V(x_1, \ldots, x_n; \alpha) = \prod_{i=1}^{n} f(x_i; \alpha), \tag{10}$$

α is mathematical expectation.

A strictly increasing transformation does not change a maximum. Maximum likelihood can also be written as:

$$\log(V(x_1, \ldots, x_n; \alpha)), \tag{11}$$

Then

$$\log(V(x_1, \ldots, x_n; \alpha)) = \sum_{i=1}^{n} \log(f(x_i; \alpha)). \tag{12}$$

– For a discrete sample:

$$f(X; \alpha) = P_\alpha(X = x_i), \tag{13}$$

$P_\alpha(X = x_i)$ represents discrete probability where α appears.

– Maximum likelihood for a discrete sample $P_\alpha(x_i)$ representing the discrete probability where α appears:

$$\log(V(x_1, \ldots, x_n; \alpha)) = \sum_{i=1}^{n} \log(P_\alpha(x_i)). \tag{14}$$

Actually, we maximize the logarithm of likelihood function to compare several models. According to (Olivier, Jouzel, El Matouat & Courtellemont 1996), principle of maximum likelihood results in over-parameterization of the model to have good performances. Penalization of likelihood value can overcome this disadvantage. Most famous penalized log-likelihood criterion is the AIC (Shang & Cavanaugh 2008), even if it is not completely satisfactory: it improves maximum likelihood principle but also led to an over-parameterization. Other traditional criteria, BIC and HQC, ensure a better estimation by penalizing oversizing models.

3.4 Akaike Information Criterion

According to (Ash 1990), entropy of a random variable is a regularity measurement. We can easily extend this concept to a model having several random variables. In their report, (Lebarbier & Mary-Huard 2004) describe all assumptions necessary to its implementation.

$$AIC = -2 \ln V + 2k, \tag{15}$$

k is the number of free parameters, $2k$ is the penalty, V is the likelihood.

The best model is the one which has the weakest AIC. This criterion uses maximum likelihood principle seen in (14). It penalizes models with too many variables, and avoids over-learning models. In the literature, Akaike Information Criterion (AIC) is often associated with another known criterion, called Bayes Information Criterion (BIC).

3.5 Bayesian Information Criterion

BIC penalizes more over-parameterized models. It was introduced in (Schwarz 1978) and is different for the correction term:

$$BIC = -2 \ln V + k \ln(n), \tag{16}$$

k is the number of free parameters of Markov Model (Avila 1996), n is the number of data, $k \ln(n)$ is the penalty term.
Like AIC, the best model is the one which gets the minimum value of BIC. Choosing between these two criteria is to choose between a predictive model and an explanatory model (Lebarbier & Mary-Huard 2004). It checks the validity of a particular model but it is mainly used to compare several models together. AIC criterion is less relevant than BIC for over-learning models.

3.6 Statical tests

Most statistical tests assume that samples are taken at random to achieve (Steinebach 2006). This sounds easy but is actually quite difficult to achieve.

3.6.1 Kolmogorov-Smirnov test

Kolmogorov-Smirnov test is a statistical test that may be used to determine if a set of data comes from a particular probability distribution ((Rukhin et al. 2010), (Bercu & Chafaï 2007)).

Empirical distribution function $F_n(x)$ for $X_1, \ldots, X_n$

sample is defined by:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i \leq x}, \quad (17)$$

$$\delta_{X_i \leq x} = \begin{cases} 1 & \text{si } X_i \leq x, \\ 0 & \text{sinon}. \end{cases}$$

The Kolmogorov-Smirnov test statistic is defined as follows:

$$D_n = \sup_x |F_n(x) - F(x)|. \quad (18)$$

3.6.2 Aspin-Welch test

Aspin-Welch's test (Welch 1951), is defined by t statistic in the following formula:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\sigma^2(\frac{1}{n_1} + \frac{1}{n_2})}}, \quad (19)$$

$$\sigma^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2 - 2}. \quad (20)$$

- $\overline{x}_i$: the i^{th} sample mean,
- σ: an estimator of the common standard deviation of the two samples,
- σ_i: samples standard deviation,
- n_i: sample size.

3.7 Epistemic uncertainties

This uncertainty is explicitly due to the design of the mathematical model. It is related to the human interpretation of the phenomenon which leads to imperfections in the design. We examine epistemic errors on our synthetic model and determine elements with the lowest uncertainty.

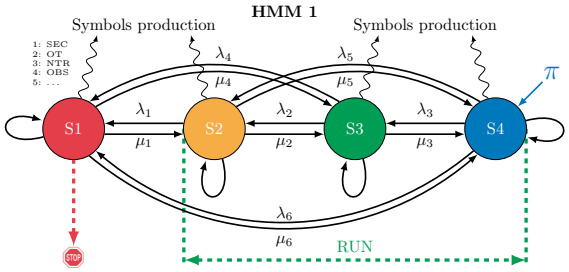
For a n measures series of $x_1, x_2, \ldots, x_i, \ldots, x_n$, the uncertainty on the average according to (Pibouleau 2010) is:

$$\Delta\overline{x} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \overline{x})^2}. \quad (21)$$

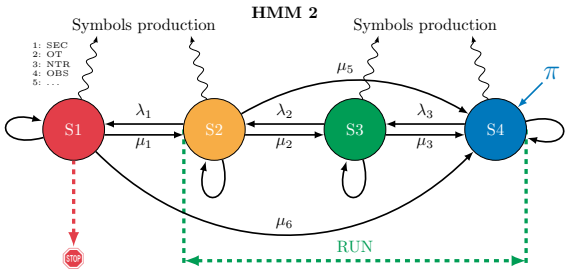
- σ: samples standard deviation.

3.8 Evaluation process

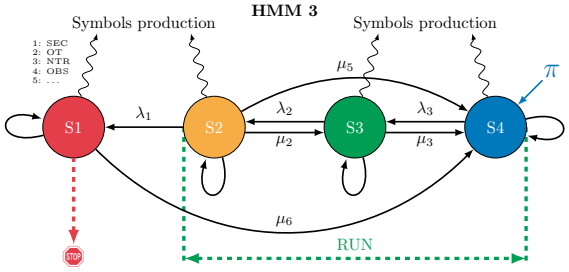
We try to evaluate the best Hidden Markov Model topologies presented in figure 2, by using all criteria shown above. Calculation is made with states and observations of three different HMM topologies (figure 2). Symbols (= i.e. observations) are produced by a synthetic HMM (the reference model), using two different learning algorithms and two different distributions of symbols.
Stochastic automata represent the degradation level of an industrial process, *S4* to *S1*, see figure 2. {S4, S3, S2} states, when process is running ("RUN"), and {S1} state, when process is stopped ("STOP"). Transitions (S4 → S3) and (S3 → S2) show progressive degradations of the process.



(a) Topology 1



(b) Topology 2



(c) Topology 3

Figure 2: Four states Hidden Markov Models.

4 EVALUATION PROCESS

We use synthetic model to produce about 1000 data events. These simulated symbols, according to real

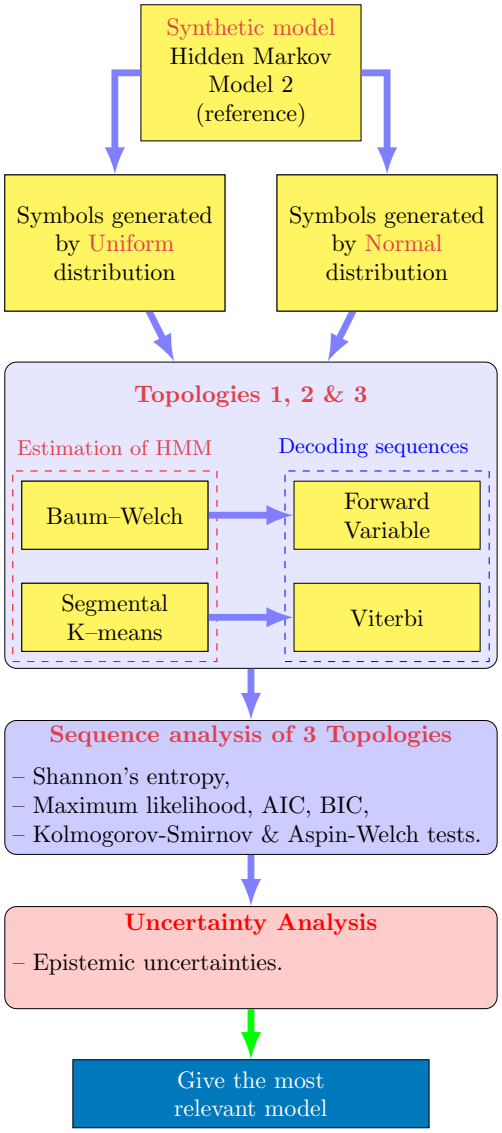


Figure 3: Matching model method, using synthetic model.

industrial process, are obtained by using uniform and normal distribution. Correlatively, we produce 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 (diagram of this process is given in figure 3): Baum-Welch learning, decoding by Forward Variable and Segmental K-means learning, decoding by Viterbi.

4.1 Simulated industrial Computerized Maintenance Management System

Nowadays, every industrial factory uses *preventive maintenance*. Maintenance agents can consign their actions and observations in a centralized database (see table 1). For example, symbols "PM, OT, SP, ..." could characterize maintenance activities

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	13/01/2007	Padlock	NTR	30	7	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.

carried out on industrial process. We recall the meaning of selected symbols resulting from observations, in table 2. "SP" symbol corresponds to a stop of production units: process state = "STOP" in table 2. It is a critical condition that our research tries to minimize. Process state = "RUN" when production units are running without failure. We study here this kind of

	Process states
	RUN
	STOP

	Interventions type	
1	SP	(Troubleshooting / Stop Production)
2	SM	(Setting Machine)
3	OT	(Other)
4	OBS	(Observation)
5	PM	(Preventive Maintenance, Production not stopped)
6	SEC	(Security)
7	PUP	(Planified Upgrading)
8	CM	(Cleaning Machine)
9	PMV	(Preventive Maintenance Visit)
10	NTR	(Nothing to report)

Table 2: Symbolic coding system of maintenance interventions.

maintenance by using *synthetic model* (§4.2) to simulate real industrial environnement. We choose "λ_i" (failure rate) and "μ_i" (repair rate) of HMM parameters (Vrignat et al. 2010), to match as possible, with maintenance recording (table 1).

4.2 Synthetic model

We make our synthetic model with Matlab by using four states oriented topology 2 presented in figure 2(b). We use this model feature because it has good performance in maintenance activities (Vrignat et al. 2010). Then, we build sequences of data (also named "signature") using this model as the reference model, by injecting stochastic symbols in this HMM. We use these symbols sequences as Markov chain (see

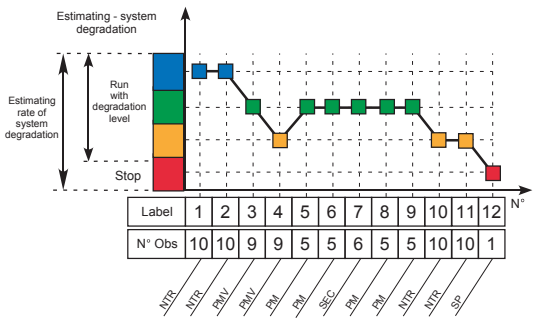


Figure 4: Degradation of process.

table 3), to model degradation level of a process (ex-

ample in figure 4).

PM	PM	SEC	PM	PM	NTR	NTR	SP	...

Table 3: Sequence of a message from maintenance database.

These simulated symbols, according to real industrial process (Vrignat et al. 2010), are obtained by using uniform and Normal (Gaussian) distribution (see figures 5 and 6). We use these symbols to train three different HMM topologies, described in figure 2, by using two different learning and decoding algorithms: Baum-Welch learning, decoding by Forward Variable, and Segmental K-means learning, decoding by Viterbi seen before.
About 1000 symbols were produced by reference model (see figures 5 and 6). Each sequence ends with a stop of production (symbol SP in red) see figure 4. We get 11 sequences in our 1000 simulated symbols. You can see distribution symbols/states for the first sequence of HMM 1: HMM 1/Baum-Welch and HMM 1/Segmental K-means algorithms, in figure 10. Finally, we obtain states sequences for each HMM outside. Later, these states are used to make comparisons between 3 different HMM topologies (figure 2), with statistical tests studied in section 3. Results are shown in section 5. Diagram of our evaluation process is given in figure 3.

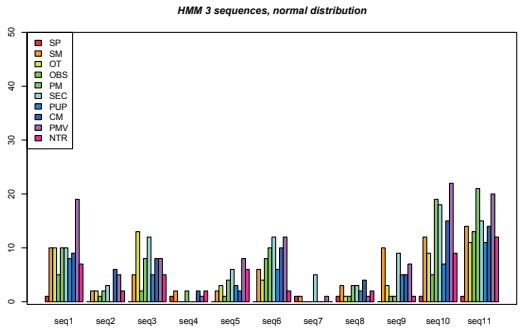


Figure 5: HMM sequences example, Normal distribution.

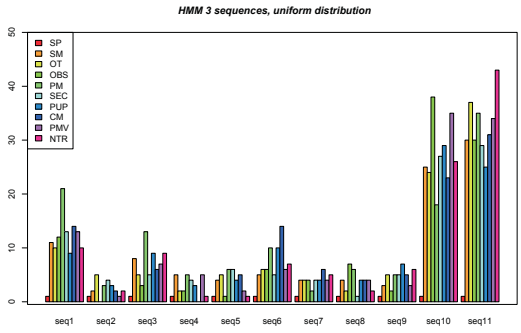


Figure 6: HMM sequences example, Uniform distribution.

5 RESULTS AND DISCUSSION

We first discuss the choice of synthetic model reference as the oriented model 2. Knowing transition probabilities of states/symbols, we tested several different topologies on different learning algorithms. At the end of the comparatives tests, we concluded that the best topology, according to failure detection, among different learning algorithms, was the topology 2 (Vrignat et al. 2010).

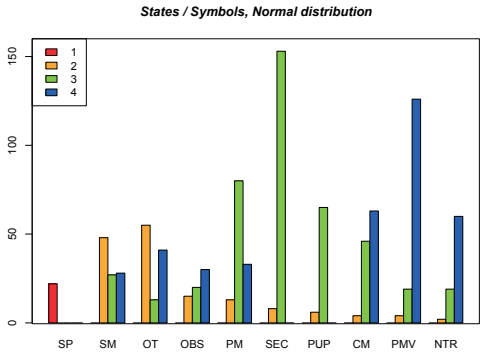


Figure 7: HMM sequences example, Normal distribution.

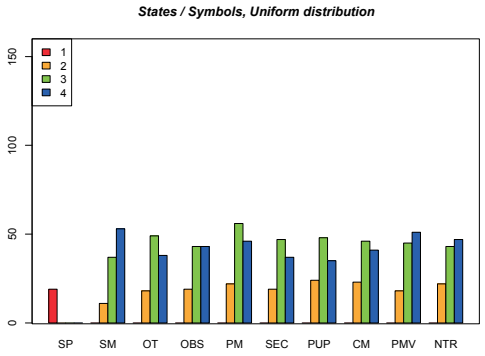


Figure 8: HMM sequences, Uniform distribution.

Afterwards, we verified the randomness of stochastic states generated by the synthetic model (figures 8 and 7). The random number generator is tested by *frequency test* § 3.1. Results in table 4, showed that all

NIST Test	P-value	
HMM	Uniform law	Normal Law
Topology 1	0.47	0.06
Topology 2	0.30	0.02
Topology 3	0.47	0.06

Table 4: p-value of states generated by synthetic model.

p-value ≥ 0.01 for all HMM figure 2. Then we could

consider that sequences of the generator are random enough to apply others tests.

Without a priori knowledge, we could give the most relevant model in the way of Shannon. Namely, we verified that the best model (which provided the better estimation of degradation level (Vrignat et al. 2010)) obtained a good "entropic" score through entropic filter, illustrated in figure 9(a). The best one was topology N°2 with Baum–Welch learning, where entropy is maximum. It also highlight the best learning algorithm recommended in (Vrignat et al. 2010): *Baum–Welch with Forward variable decoding*, with *normal distribution* of symbols.

Afterwards, we evaluated likelihood (or probability) of observations sequences given by synthetic HMM. Results of maximum likelihood and *BIC* highlight the most relevant topology: *HMM 2*, fig 2(b). That corroborates (Vrignat et al. 2010) results. On the other side, our results did not show clearly, differences between algorithms, we could not conclude for the best learning and decoding algorithm. Nevertheless, with Segmental K–means algorithm, in figure 10(c), the reader can see a bad distribution of symbols. *AIC* does not penalize our 1000 data, thats why *BIC* is more suitable, because of "$k \ln(n)$" term of equation 16.

Next, we applied statistical tests on our three HMM topologies. Aspin-Welch and Kolmogorov-Smirnov figure 9(b) give the same results: most relevant model is *topology 2*, *Baum-Welch learning algorithm* with Forward Variable decoding is the best learning algorithm and finally, stochastic symbols generated with *normal distribution* is the best one.

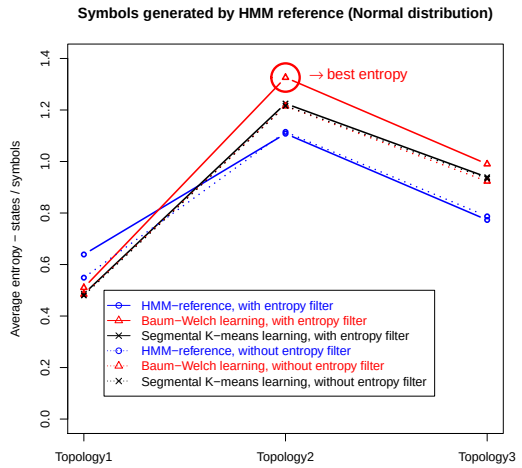
Finally, epistemic uncertainties depicted in figure 9(c), highlight once more that *topology 2*, *Baum-Welch learning algorithm* with Forward Variable decoding and *Normal* distribution gives the lowest error rate. Reader can find all results in table 5. Unfortunately, we failed to establish any ranking between these various criteria.

	Topology			Learn algo		Distribution	
Evaluation criteria	**1**	**2**	**3**	**BW**	**SK**	**Nor.**	**Uni.**
Shannon's Entropy		×		×		×	
Maximum likelihood		×		**No finding**		**No finding**	
AIC		×		**No finding**		**No finding**	
BIC		×		**No finding**		**No finding**	
Aspin-welch test		×		×		×	
Kolmogorov-Smirnov		×		×		×	
Best uncertainty		×		×		×	

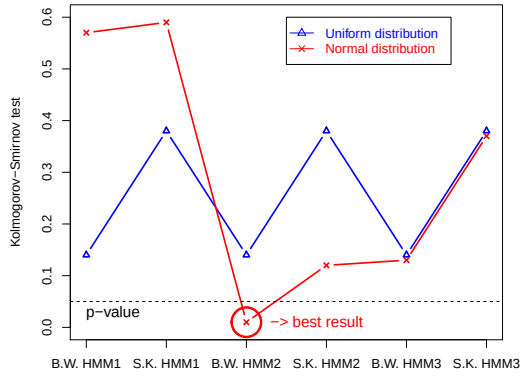
Table 5: General results for some criteria.

6 CONCLUSION

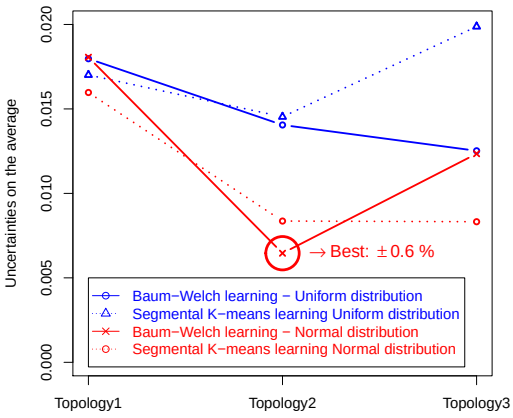
After testing randomness of our synthetic model generator, we have applied all criteria studied above on three different HMM topologies. We have successfully applied this method to three different models. The first one, uses *Shannon's entropy* and entropic filter. Given a set of observations sequences simulated by our synthetic model, we verified that the



(a) Average entropy, Gaussian distribution



(b) Kolmogorov-Smirnov test statistic



(c) Stochastic uncertainties

Figure 9: Evaluation results.

most relevant model obtained a good "entropic" score. That corroborates (Vrignat et al. 2010) results which showed that *topology N°2* was the one which came closest to real industrial process. This criterion also showed that *Baum–Welch* learning algorithm with *Forward Variable* decoding gave best results. Moreover Shannon's entropy showed that *Normal distribution* was the best one to simulate industrial observations. Maximum likelihood and *BIC* emphasized that HMM 2 was the best topology. Unfortunately, these criteria were too near each other to make conclusions about learning algorithm. With Aspin-Welch and Kolmogorov-Smirnov test, we could verify that the most relevant model had the *"goodness of fit"* i.e. how well model fits the set of observations sequences. The statistical way told us the same conclusions than entropic results (topology, learning algorithm and distribution). Same goes for errors of epistemic uncertainties: *topology N°2*, *Baum-Welch* learning algorithm with *Forward Variable* decoding and *normal distribution* of stochastic symbols gave the *lowest error rate*. Thus, we specified our analysis from (Roblès, Avila, Duculty, Vrignat & Kratz 2011) paper.
We gave heterogeneous methods to help expert of maintenance to choose and select the best way to optimize his maintenance policy. Indeed, when HMM output will indicate an **orange level** (S2), the expert would decide for a preventive maintenance before the breakdown. Good relevance and good errors rate for topology N°2, Baum-Welch algorithm / decoding by Forward Variable and a Gaussian distribution of observation sequences, allow us to apply these results as part of preventive maintenance applications. Indeed, in our work on industrial breakdown prediction, determining the best model is expected to reduce significantly failure rate in production. Minimizing failure rate, will reduce dangerous human intervention in maintenance, especially in an unsafe working environment. Decreasing machines failures will furthermore reduce power consumption and thus reduce release of CO_2.

In further work, we will try to test robustness of our synthetic model with different noises. Our research goals are to validate a real choice of a model: topology, symbol,... without a priori knowledge on results.

References

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

Ash, R. (1990). Information theory, *Dover Publications* .

Avila, M. (1996). *Optimisation de modèles Markoviens pour la reconnaissance de l'écrit*, PhD thesis, Université de Rouen.

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

Bercu, B. & Chafaï, D. (2007). *Modélisation stochastique et simulation - Cours et applications*, Collection Sciences Sup - Mathématiques appliquées pour le Master, Société de Mathématiques Appliquées et Industrielles (SMAI), éditions Dunod.

Chandrasekaran, V., Johnson, J. K. & 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* .

Chen, S. S. & Gopalakrishnan, P. S. (1998). Speaker, environment and channel change detection and clustering via the bayesian information criterion, *Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop*, Lansdowne, Virginia, USA.

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

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

Hocker, D., Xiaohu, L. & Iyengar, S. S. (2011). Shannon entropy based time-dependent deterministic sampling for efficient on-the-fly quantum dynamics and electronic structure, *J. Chem. Theory Comput.* pp. 256–268.

Juang, B. H. & 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).

Lebarbier, E. & Mary-Huard, T. (2004). Le critère bic : fondements théoriques et interprétation, *Research Report RR-5315*, INRIA.

Olivier, C., Jouzel, F., El Matouat, A. & Courtellemont, P. (1996). Prediction with vague prior knowledge, *Communications in Statistics 25-Theory and Methods* pp. 601–608.

Pibouleau, L. (2010). *Assimiler et utiliser les statistiques*, Ellipses Marketing, technosup.

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning)*, 1 edn, Morgan Kaufmann.

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* pp. 257–286.

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

Rukhin, A., Soto, J., Nechvatal, J., Barker, E., Leigh, S., Levenson, M., Banks, D., Heckert, A., Dray, J., Vo, S., Rukhin, A., Soto, J., Smid, M., Leigh, S., Vangel, M., Heckert, A., Dray, J. & Iii, L. E. B. (2010). A statistical test suite for random and pseudorandom number generators for cryptographic applications.

Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics* **6**: 461–464.

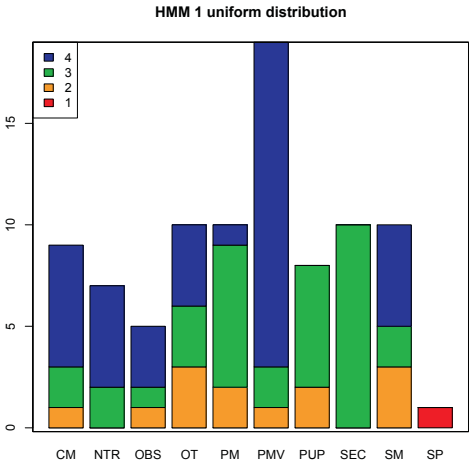
Shang, J. & Cavanaugh, J. E. (2008). Bootstrap variants of the akaike information criterion for mixed model selection, *Comput. Stat. Data Anal.* **52**: 2004–2021.

Steinebach, J. (2006). E. l. lehmann, j. p. romano: Testing statistical hypotheses, *Metrika* **64**: 255–256.

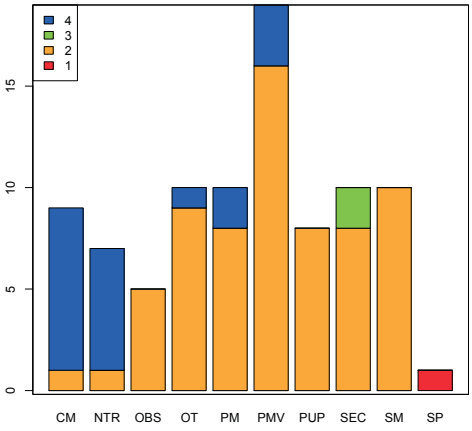
Vrignat, P., Avila, M., Duculty, F., Aupetit, S., Slimane, M. & Kratz, F. (2011). Maintenance policy: degradation laws versus hidden markov model availability indicator, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* .

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

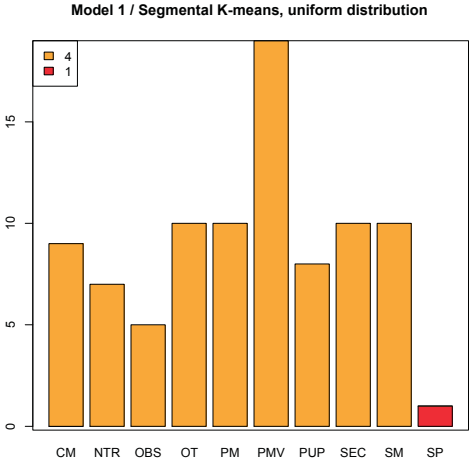
Welch, B. L. (1951). Welch's k-sample test, *Biometrika* **38**: 330–336.



(a) HMM–Reference



(b) Baum–Welch



(c) Segmental K–means

Figure 10: First sequence, using normal distribution.