

Hidden Markov Models in Dynamic System Modelling and Diagnosis

Tarik Al-ani

¹Laboratoire des Systèmes d'Ingénierie de Versailles, Université de Versailles
Saint-Quentin, Centre Universitaire de Technologie, 10/12 Avenue de l'Europe - 78140

Vélizy, France

²Département of Informatique, ESIEE-Paris, Cité Descartes-BP 99 93162 Noisy-Le-Grand
France

1. Introduction

Over the past few decades, Hidden Markov models *HMMs* have been widely applied as a *data-driven modeling* approach in automatic speech recognition (Rabiner, 1989). In this field, signals are encoded as temporal variation of a short time power spectrum. *HMMs* applications are now being extended to many fields such as pattern recognition (Fink, 2008); signal processing (Vaseghi, 2006); telecommunication (Hirsch, 2001); bioinformatics (Baldi et al., 2001), just to name a few applications.

Recently, *HMMs* represent in a natural way the individual component states of a dynamic system. This fact make them useful in biomedical signal analysis and in medical diagnosis (Al-ani et al., 2004; 2008; Al-ani & Trad, 2010a; Daidone et al., 2006; Helmy et al., 2008; Novák et al., 2004a;b;c;d). For the same reason, they are used in fault detection and mechanical system monitoring (Al-ani & Hamam, 2006; Bunks et al., 2000; Heck & McClellan, 1991; Miao et al., 2007; Smyth, 1994) as well in modelling, identification and control of dynamic systems (Elliot et al., 2004; Frankel, 2003; Fraser, 2010; Kwon1 et al., 2006; Myers et al., 1992; Tsontzos et al., 2007; Wren et al., 2000). An *HMM* may be used also to describe discrete stochastic changes in the system and then folds in the continuous dynamics, by associating a set of dynamic and algebraic equations with each *HMM* model¹. This leads to a model called *probabilistic hybrid automaton (PHA)* for short *hybrid complex Systems* (Hofbauer, 2005).

In these fields, it is essential to effectively learn all model parameters from a large amount of training data according to certain training criteria. It has been shown that success of *HMMs* highly depends on the goodness of estimated models and the underlying modeling technique plays an critical role in the final system performance.

The objective of this chapter is to sensitize the reader on two problems related to conventional *HMMs* which are important for dynamic system modelling and diagnosis: the training problem and the data-driven selection methods of the *structure* for the constructed *HMM* and then to introduce two application examples of *HMMs* in diagnosis of mechanical and medical dynamic systems.

2. Hidden Markov Models (HMMs)

This section introduces briefly the mathematical definition of Hidden Markov Models. We introduce only their conventional training aspects. The notations will be done to remain in the contexts cited by Rabiner (Rabiner, 1989).

The HMMs are double stochastic processes with one underlying process (state sequence) that is not observable but may be estimated through a set of processes that produce a sequence of observations. HMMs are a dominant technique for sequence analysis and they owe their success to the existence of many efficient and reliable algorithms.

Consider a discrete time Markov chain with a finite set of states $S = \{s_1, s_2, \dots, s_N\}$. A HMM is defined by the following compact notation to indicate the complete parameter set of the model $\lambda = (\mathbf{\Pi}, \mathbf{A}, \mathbf{B})$ where $\mathbf{\Pi}$, \mathbf{A} and \mathbf{B} are the initial state distribution vector, matrix of state transition probabilities and the set of the observation probability distribution in each state, respectively (Rabiner, 1989).

$$\mathbf{\Pi} = [\pi_1, \pi_2, \dots, \pi_N], \pi_i = P(q_1 = s_i), \mathbf{A} = [a_{ij}], a_{ij} = P(q_{t+1} = s_j | q_t = s_i),$$

$1 \leq i, j \leq N, s_i, s_j \in S, t \in \{1, 2, \dots, T\}$. In the rest of this chapter, the states s_i and s_j will be written as i and j respectively for simplicity.

The observation at time t , \mathbf{O}_t , may be a discrete symbol (*Discrete HMMs (DHMMs)* case), $\mathbf{O}_t = v_k, v_k \in V = \{v_1, v_2, \dots, v_M\}$, or continuous, $\mathbf{O}_t \in \mathbb{R}^K$. For a discrete observation, v_k will be written as z for simplicity.

The observation matrix \mathbf{B} is defined by $\mathbf{B} = [b_i(\mathbf{O}_t)]$, where $b_i(\mathbf{O}_t)$ is the state conditional probability of the observation \mathbf{O}_t defined by $b_i(\mathbf{O}_t) = P(\mathbf{O}_t = z | q_t = i), 1 \leq i \leq N, 1 \leq z \leq M$. For a continuous observation (*Continuous HMMs (CHMMs)* case), $b_i(\mathbf{O}_t)$ is defined by a finite mixture of any log-concave or elliptically symmetric probability density function (*pdf*), e.g. Gaussian *pdf*, with state conditional observation mean vector μ_i and state conditional observation covariance matrix Σ_i , so \mathbf{B} may be defined as $\mathbf{B} = \{\mu_i, \Sigma_i\}, i = 1, 2, \dots, N$. The model parameters constraints for $1 \leq i, j \leq N$ are

$$\sum_{i=1}^N \pi_i = 1, \sum_{j=1}^N a_{ij} = 1, a_{ij} \geq 0, \sum_{k=1}^M b_i(\mathbf{O}_t = z) = 1 \text{ or } \int_{-\infty}^{\infty} b_i(\mathbf{O}_t) d\mathbf{O}_t = 1.$$

In the case of *left-to-right HMMs* (Rabiner, 1989), multiple observation sequences must be used. Unlike the *ergodic HMM* case, the *left-to-right HMM* necessitates taking into account that the initial state probability is always equal to one. Thus, all the sequences are supposed to start with the same state. In this work, the specific structures are taken into account by weighting the values of the elements of the matrix \mathbf{A} . The weights are defined by zeros and ones elements corresponding to the structure of the matrix \mathbf{A} .

In general, at each instant of time t , the model is in one of the states $i, 1 \leq i \leq N$. It outputs \mathbf{O}_t according to a discrete probability (in the *DHMMs* case) or according to a continuous density function (in the *CHMM* case) $b_j(\mathbf{O}_t)$ and then jumps to state $j, 1 \leq j \leq N$ with probability a_{ij} . The state transition matrix defines the structure of the HMM (Rabiner, 1989). The model λ may be obtained off-line using some training algorithm. In practice, given the observation sequence $O = \{\mathbf{O}_1 \mathbf{O}_2 \dots \mathbf{O}_T\}$, and a model λ , the HMMs need three fundamental problems to be solved. Problem 1 is how to calculate the likelihood $P(O|\lambda)$? The solution to this problem provides a score of how O belongs to the model λ . Problem 2 is how to determine the most likely state sequence that corresponds to O ? The solution to this problem provides the sequence of the hidden states corresponding to the given observation sequence O . Problem

3 is how to adjust the model λ in order to maximize $P(O|\lambda)$? This is the problem of estimating the model parameters given a corpus of training observations sequences. Problems 1 and 2 are solved in the decoding stage using either the forward probability (Rabiner, 1989) or the *Viterbi decoding algorithm* (Viterbi, 1967), while problem 3 is solved during the training stage using either a conventional algorithm such as the *Baum-Welch* algorithm and the Viterbi-based algorithm, known as *segmental K-mean* in the case of *CHMMs* (Rabiner et al., 1986; Rabiner, 1989) or some other new training algorithm.

2.1 The decoding problem

In some applications, for a given *HMM*, the most probable state sequence $Q^* = \{q_1^* q_2^* \dots q_T^*\}$ which generates a given observation sequence $O = \{O_1 O_2 \dots O_T\}$ could be desired. To estimate the optimum state sequence, the stochastic dynamic programming based *Viterbi algorithm* may be used. The Viterbi algorithm finds the state sequence Q^* such that

$$Q^* = \arg \max_Q P(Q|O, \lambda) = \arg \max_Q P(O, Q|\lambda)$$

We define the probability function $\delta_t(i)$ at the i th state, time t of the multi-stage by

$$\delta_t(i) = \max_{q_1 q_2 \dots q_t} \ln p(q_1 q_2 \dots, q_t = i, O_1 O_2 \dots O_t | \lambda, q_t),$$

$t \geq 1$, i.e. the maximum partial observation probability associated with state sequence of *HMM* at time t . In order to allow for path tracking, $\Psi_t(i)$ is also introduced to keep track of the most probable state sequence visited by an observation O_t .

2.2 The training problem

The *HMMs* owe their success to the existence of many efficient and reliable *generative training* algorithms. The *generative training* has the advantage that one may impose all kinds of structure constraints, such as graphical models (Jordan, 2004). These constraints are related to inherent dependency or various relationship of data. More recently, a hybrid *generative/discriminative training* schemes were proposed for training. In a more general sense, the *discriminative training* of *generative models* may include any alternative estimation methods for traditional *generative models* based on a different training criterion rather than the *maximum likelihood estimation (MLE)*. The *discriminative training* directly optimises a *mapping function* from the input data samples to the desired output labels by adjusting a decision boundary in the feature space. Several *mapping functions* can be estimated using some criteria that are directly relevant to the ultimate classification and regression purpose. The more used *mapping functions* in the literature are the *conditional maximum likelihood (CML)* (Jebara & Pentland, 1998), also known as *maximum mutual information (MMI)* estimation (Bahl et al., 1986; Woodland & Povey, 2002), empirical risk minimization (*ERM*) (Meir, 1995) and *large margin estimation (LME)* (Scholkopf & Smola, 2002; Smola et al., 2000).

2.2.1 Conventional Baum-Welch training

In many conventional *generative training* methods, a new set of parameters $\tilde{\lambda}$ is chosen such that $P(O|\tilde{\lambda})$ is maximized for a given observation sequence $O = \{O_1 O_2 \dots O_T\}$. In these methods, a new set of parameters $\tilde{\lambda}$ is chosen such that $P(O|\tilde{\lambda})$ is maximized for a given observation sequence $O_{t+1} O_{t+2} \dots O_T$. The re-estimation formulas of a given *HMM* (Rabiner,

1989) may be derived directly by maximizing (using standard constrained optimization techniques) the *Baum's auxiliary function* (Baum et al., 1970),

$$Q_{aux}(\lambda, \tilde{\lambda}) = \sum_Q P(O, Q|\lambda) \log P(O, Q|\tilde{\lambda}),$$

over $\tilde{\lambda}$. It has been proven by Baum-Welch & Eagon (Baum & Eagon, 1967) that maximization of $Q_{aux}(\lambda, \tilde{\lambda})$ leads to increased likelihood, i.e.

$$\max_{\tilde{\lambda}} [Q_{aux}(\lambda, \tilde{\lambda})] \Rightarrow P(O|\tilde{\lambda}) \geq P(O|\lambda).$$

Eventually the likelihood function converges to a critical point.

In the literature, the solution space of *HMMs* is usually coded as a function of the model parameters $(\mathbf{\Pi}, \mathbf{A}, \mathbf{B})$. The *MLE* or the probability of the observation sequence associated to the *HMM* parameters may be applied. This may be obtained using the *Expectation-Maximization* algorithm (*EM*) (Dunmur & Titterton, 1998). This algorithm leads to *Baum-Welch reestimation formulas* which are based on the calculation of the forward and backward probabilities (Rabiner, 1989). The *EM* algorithm is slow, computationally expensive in practice and its estimation performance in the case of *CHMMs* depends on the given initial model.

The *Baum-Welsh* algorithm has become very popular due to its simplicity and efficiency. However, this algorithm may be easily trapped in a local optimum.

2.2.2 Conventional *K-mean* (Viterbi-based) training

Viterbi training (or *segmental K-means*) method is proposed as a *generative training* approach to speed up the learning process. The basic idea behind *Viterbi training* is to replace the computationally costly expectation E-step of the *EM* algorithm by an appropriate maximization step with fewer and simpler computations. Thus the *MLE* criterion can be approximated by maximizing the probability of the best *HMM* state sequence for each training sample, given the model. The most probable path for each training sequence is derived using *Viterbi decoding algorithm*. Based on this path, the state transition and observation emissions occurrences are estimated and used for the reestimation of the *HMM* parameters:

- Initialization: Choose model parameters randomly.
- Iteration:
 - Derive the most probable state sequence Q using the *Viterbi decoding algorithm*.
 - Calculate a_{ij} and $b_i(\mathbf{O}_t)$ for the given Q .
 - Estimate the new model parameters using the estimated occurrences of state transition and observation emissions.
- Termination: Stop, if convergence of the model parameters.

As mentioned by Rodríguez & Torres (Rodríguez & Torres, 2003) and by Koloydenko (Koloydenko, 2007), *Viterbi training* involves much less computational effort than *Baum-Welch*, still providing the same or slightly worse performance, so it is a common choice among designers of speech recognition systems.

2.2.3 New generative and discriminative training algorithms

There is no theoretical method to overcome the model initialisation and the local optimum problems. In practice, many *generative training* as well as *discriminative training* methods were proposed to overcome the limitations of the above conventional algorithms.

2.2.3.1 New generative training algorithms for the generative models

Some authors (Kanevsky, 2004; Povey & Kingsbury, 2007; Woodland & Povey, 2002) suggested an *Extended Baum-Welch (EBW)* algorithm which is an iterative algorithm that uses a specific set of update rules performed at each iteration. Huda et al. (Huda et al., 2006) proposed different initial guesses and the solution that corresponds to the local maximum with the largest probability is selected. The authors (Aupetit et al., 2007; Fengqin et al., 2008; Xue et al., 2006) used the *Particle Swarm Optimization (PSO)* algorithm (Kennedy et al., 2001). Wang et al. (Wang et al., 2008) proposed an HMM based on a *mixture of factor analysis (HMM-MFA)* to model the correlation between the feature vector elements in speech signals. Turner (Turner, 2008) proposed an effective implementation of a direct approach using an adaptation of the *Levenberg-Marquardt* algorithm. This approach is based on the possibility of maximizing the likelihood of HMMs by means of a direct *gradient-Hessian* approach, without resorting to the EM algorithm.

Meta-heuristics approaches such as *simulated annealing* or *Tabu search* algorithms (Aarts & Korst, 1990; Kinney et al., 2007; Kirkpatrick et al., 1983; Lee, 2008; Lyu et al., 1996; Mazumdar et al., 2007; McKendall & Shang, 2008) were used to minimize general energy or cost functions with the aim of finding global optimal solutions within the feasible parameter domain. Due to the non convexity of the *maximum likelihood criterion*, some meta-heuristics such as evolutionary programming and simulated annealing algorithms in their original implementation in the context of HMMs have been proposed (Andrieu et al., 2000; Eddy, 1995; Lee & Park, 2007; Paul, 1985; Pérez et al., 2007; Rao & Rose, 2001; Tang et al., 2006; Yang et al., 2008; Won et al., 2007; Zhao et al., 2007). In these works, the *simulated annealing* algorithm is applied to the CHMM parameters. It is well known in *simulated annealing* that the convergence and the size of the chain used at each temperature depend on the size of the solution space. A continuous coding of the solution space leads to an infinite size of this space which gives rise to convergence difficulties whereas the discrete coding of the solution space leads to a finite size of this space which gives rise to better convergence properties (Al-ani & Hamam, 2010b).

2.2.3.2 New discriminative training algorithms for the generative models

Based on the *mapping functions*, many *discriminative models* were developed (Jiang, 2010). Particularly, based on the *generalization bounds* in *statistical learning theory* (Vapnik, 1998), *large margin estimation (LME)* (Scholkopf & Smola, 2002; Smola et al., 2000) have been applied with success in several fields. However, as stated by Jiang (Jiang, 2010), there are still some limitations in the *discriminative training* scheme. For example, it is not straightforward to deal with *latent variables* and exploit the underlying structure of data in *discriminative models*. Moreover, computational complexity is considerably higher in discriminative training since it requires simultaneous consideration of data from all classes. Hence, no standalone discriminative model can yield comparable performance as *generative models*, i.e., HMMs, on any significant application task.

More recently, several hybrid algorithms have emerged combining *generative models* and *discriminative training* based on some *discriminative criteria* that are related to the application purposes (Altun et al., 2003; Jaakkola & Haussler, 1998; Jaakkola et al., 1999; Taskar et al., 2003).

This scheme makes possible to deal with latent variables and exploit the underlying structure of data. Discriminative training of *HMMs* have been applied in many fields:

- *Speech and handwritten recognition* (Altun et al., 2003; Arslan & Hansen, 1999; Bach & Jordan, 2005; Cheng et al., 2009; 2010; Do & Artières, 2009; Dong & Li, 2007; Ganapathiraju et al., 2000; Golowich & Sun, 2007; Gu et al., 2000; Jiang et al., 2006; Jiang & Li, 2006; Jiang, 2010; Krüger et al., 2005; 2006; Lee & Park, 2005; Li & Jiang, 2005; Li et al., 2005a;b; Li & Juang, 2006; Liu et al., 2007a;b; Markov et al., 2001; Sanguansat et al., 2004; Saon & Povey, 2008; Sha & Saul, 2009; Solin & Burshtein, 2008),
- *Computational biology* (Li & Noble, 2003),
- *Dialog act tagging* (Surendran & Levow, 2006),
- *Analysis of Facial expression temporal dynamics* (Valstar & Pantic, 2008)
- *Financial applications* (Rao & Hong, 2001),
- *Information extraction* (Altun et al., 2003),
- *Modeling chaotic nonlinear dynamical systems* (Myers et al., 1992),
- *Teleoperation tasks* (Castellani et al., 2004),
- *Temporal signals and EEG analysis* (Xu et al., 2005; 2006),

2.3 Selection of the number of states in *HMMs*

A practical implementation but fundamental issue to be addressed when using *HMMs* is the determination of their *structure*, namely the *topology* (or the transitions between the states) and the number of states. The *structure* of an *HMM* is determined by some constraints that may be introduced in the *HMM structure*, such as forcing the presence or absence of connections between certain states.

Several approaches were proposed for learning the *structure* of *HMMs*. Stolcke & Omohundro (Stolcke & Omohundro, 1993) proposed a data-driven technique where the induction process starts with the most specific model consistent with the training data and generalises by successively merging states. Both the choice of states to merge and the stopping criterion are guided by the Bayesian posterior probability. Takara et al. (Takara et al., 2007) proposed the application of a *genetic algorithm* (*GA*) to search out an optimal *structure*. Biem et al. (Biem et al., 2002; Biem, 2003) proposed a model selection criterion for classification problems. The criterion focuses on selecting models that are discriminant instead of models based on the Occam's razor principle of parsimony between accurate modeling and complexity. The criterion, dubbed *Discriminative Information Criterion* (*DIC*), was applied to the optimisation of *HMM topology* aimed at the recognition of cursively-handwritten digits. The results showed that *DIC* generated models achieve a good relative improvement in performance from a baseline system generated by the *Bayesian Information Criterion* (*BIC*). Bcego et al. (Bcego et al., 2003) proposed a technique which is able to deal with drawbacks of standard general purpose methods, like those based on the *Bayesian inference criterion* (*BIC*), i.e., computational requirements, and sensitivity to initialisation of the training procedures. The basic idea is to perform "decreasing" learning, starting each training how to learn model *structure* from data and how to make the best use of labeled and unlabeled data. They showed that a manually-constructed model that contains multiple states per extraction field outperforms a model with one state per field, and discussed strategies for learning the model *structure* automatically from data. They also demonstrated that the use of distantly-labeled data to set model parameters provides a significant improvement in extraction accuracy.

Siddiqi et al. (Siddiqi et al., 2007) proposed a new *state-splitting algorithm* that addresses two problems: choosing the *structure* (model selection) and estimating model parameters (learning).

Other approaches were also proposed for estimating the number of states. Celeux & Durand (Celeux & Durand, 2008) proposed a technique for selecting *HMM* state number with cross-validated Likelihood and they gave two approaches to compute cross-validated likelihood for a *HMM*. The first one consists of using a deterministic half-sampling procedure, and the second one consists of an adaptation of the *EM algorithm* for *HMMs*, to take into account randomly missing values induced by cross-validation. Gunter & Bunke (Gunter & Bunke, 2003) examined some optimization strategies for an *HMM* classifier that works with continuous feature values and uses the Baum-Welch training algorithm. The free parameters of the optimization procedure introduced in their paper are the number of states of a model, the number of training iterations, and the number of Gaussian mixtures for each state. The proposed optimization strategies are evaluated in the context of a handwritten word recognition task.

3. Some selected applications of *Hidden Markov models* in dynamic systems

In this section, we introduce two application examples on dynamic processes. The first application concerns the *condition-based maintenance* of machines using *HMMs* and the second one concerns the diagnosis in medicine. For more details, see (Al-ani et al., 2004; Bunks et al., 2000)

3.1 *Condition-based maintenance* of machines using *HMMs*

The term *condition-based maintenance* (*CBM*) is used to signify the monitoring of machines for the purpose of *diagnostics* and *prognostics*. *Diagnostics* are used to determine the current health status of a machine's internal components and *prognostics* are used to predict their remaining useful life. *CBM* has the potential to greatly reduce costs by helping to avoid catastrophic failures (an extremely important point, for example, for helicopter gearboxes) and by more efficiently determining the intervals required for maintenance schedules. The economic ramifications of *CBM* are many fold since they affect labour requirements, replacement part costs, and the logistics of scheduling routine maintenance. The reduction of maintenance to that which is strictly necessary can have the effect of prolonging machine life and that of diminishing the defects that can be introduced by the maintenance itself. Finally, the reduction of catastrophic failures which lead to the loss of life and equipment can have an important effect on insurance premiums.

One method for performing *CBM* is by using vibration measurements. The main objective is the detection of vibrational characteristics which correspond to physical changes in the machine which indicate abnormal operation. Examples of this could be chipping in a roller bearing or spalling on a pinion gear. The primary challenge is to achieve a high degree of precision in classifying a machine's health given that its vibrational characteristics will vary with many factors not all corresponding to defective components. For example, two identical, new machines will generally have different vibrational characteristics due to differences in the manufacturing process. Furthermore, a machine's underlying vibrational character is likely to change over time as a function of the operating conditions, the maintenance schedules it has undergone, and aging of the machine. Finally, the specific vibrational characteristics of the machine will change as torque loads vary. Clearly, it is important to be able to differentiate between vibrational changes which are due to machine component defects and

those due to changing operating conditions. Assuming that torque loads vary slowly with time, recorded vibration data should demonstrate practically stationary statistics over short intervals. Under these conditions the following approach suggests itself. First, all the different operating conditions which give rise to significantly different statistics in vibration data must be identified. Second, the statistics of vibration data for various defects must be determined either experimentally or by modelling. The combined sets of statistics would then serve as a universe of models with which hypothesis testing could be performed on segments of the data. Continuous processing of sequential segments of the data would result in classification of operating condition and defect type (if any) as a function of time. Prior knowledge of how a machine were to be operated or information about the relative frequency of the occurrence of different types of defects could also be used to improve the performance of the above classification algorithm. This would be accomplished by constructing a doubly stochastic process which describes probabilistically how the machine is likely to undergo transition from state to state and what the statistics are at each state. This type of process is well suited to the statistical modelling methodology defined by *HMMs*.

HMMs are well suited to modelling of quasi-stationary signals and thus, as will be discussed in what follows, can be used to perform detection and estimation for machine *diagnostics* and *prognostics*. Two features of *HMMs* are particularly useful in the monitoring of machine health. The first is that computationally efficient methods exist for computing likelihoods using *HMMs* (Van Trees, 1968). This feature is important since it promises that signal processing tools based on *HMMs* can be implemented cost effectively. Furthermore, there exist efficient techniques which can be used for system identification with *HMMs*. This means that *HMMs* can be used to build data-driven models of machines relieving somewhat the need to identify specific features in data to be used as health indicators.

Bunks et al. (Bunks et al., 2000) compared problems of speech processing, an area where *HMMs* have been applied extensively, to those of machine health monitoring. Their comparison was useful since it helped motivate the use of *HMMs* for *CBM* as well as indicated what some of the critical issues. Then, they used the Westland helicopter gearbox data to illustrate some of the characteristics of vibration data under different operating conditions and types of defects and to illustrate the application of *HMMs* to *CBM*.

3.1.1 The Westland helicopter gearbox data

The laboratory data set is made available by the Office of Naval Research for the evaluation of machine health monitoring algorithms. This data set consists of vibration measurements from a set of accelerometers placed on the casing of a Westland helicopter gearbox (Bunks et al., 2000). The gearbox was operated at various torque levels and with various seeded defects in a special test rig. This data set is used in examples of the application of *HMMs* to *CBM*.

The data set consists of 68 distinct operating conditions constituting nine different torque levels and eight different seeded defects (of which one is actually a no-defect case). For each operating condition time samples at a sampling rate of 100 kHz from eight accelerometers are available for analysis. Bunks et al. (Bunks et al., 2000) noted the following observations on the salient features of the Westland helicopter gearbox data set:

- The data are not stationary as a function of operating torque level.
- There are clear, although complex, differences in the spectra as a function of the types of defects.
- There are clear, although complex, differences in the spectra as a function of the severity of defect level.

These observations were important since they validated the comparisons made by Bunks et al. (Bunks et al., 2000), imply the feasibility of *prognostics*, and motivate the construction of *HMMs* models for *CBM*.

3.1.2 HMM modelling for the Westland helicopter gearbox data

The objective is to construct some HMM models which can be used on data from the Westland helicopter gearbox. *HMMs* are composed of two main components. There is the state transition matrix which probabilistically describes how the model moves from one stationary type of behaviour to another. There is also the probability distribution associated to each state which describes the statistics of the particular stationary model. In order to effectively model the Westland helicopter gearbox data presented briefly in the previous section it is necessary to construct a reasonable model for both of these components. This is achieved by first specifying the total number of states for the model and then by estimating the parameters of an appropriate probability density for each state. As for the state transition matrix this information can only be obtained by using a prior experimental knowledge of the frequency of occurrence of each defect and the average time spent operating at each of the torque levels. For the purposes of their work this information was not available; nevertheless, some effort was made Bunks et al. (Bunks et al., 2000) to construct a reasonable state transition model.

3.1.2.1 Modelling state probability densities

As noted in Section 3.1.1 the characteristics of the power spectra of the Westland data change significantly for different torque levels and for different defect types. For the Westland data set there are a total of 68 different torque-level defect-type pairs available. Thus, the *HMM* constructed contains the same number of states.

An essential part of applying *HMM* technology to any problem is to decide what the appropriate observations are. In general, the observations could be the raw data or some function or transformation of the data. A transformation of the data is preferable when the result allows for the diminution of the quantity of data which needs to be processed without losing the ability to effectively monitor the *HMM* process. This may be important for the Westland data since it is acquired at a rate of 100k samples per second for each of the eight sensors and thus represents a relatively heavy data volume.

Each state of the Markov chain associated to an *HMM* must have a state process model which implicitly or explicitly specifies a probability density function. In the *HMM* literature the Gaussian distribution is often used although multi-modal distributions are also used by taking mixtures of Gaussians (Liporace, 1982; Rabiner, 1989). Other common choices include mixtures of autoregressive and autoregressive moving-average models (Poritz, 1982). In this application a simple multi-dimensional Gaussian distribution was used. A different Gaussian is estimated for each of the 68 operating conditions. Each Gaussian is eight-dimensional (8D) (due to the eight sensors) and is estimated using the first 10k samples of each of the operating condition runs (i.e. the first 0.1 s of data). The mean vectors, μ_i , and covariance matrices, Σ_i for each of the $i = 1, 2, \dots, 68$ cases were obtained using the following formulas:

$$\mu_i = \frac{1}{T} \sum_{t=1}^T y_i(t) \quad (1)$$

$$\Sigma_i = \frac{1}{T} \sum_{t=1}^T [y_i(t) - \mu_i][y_i(t) - \mu_i]^T \quad (2)$$

where $T=10000$ and $y_i(t)$ is the 8-vector of observations at time $t\Delta t$ for data from operating condition i .

We note that, in principle, the vibration data should be zero mean since having a DC level of vibration is meaningless. Nevertheless, it is important to take the sample mean into account to cancel the effects of bias coming from electronic amplification. We further note that modelling the time sample data as an independent and identically distributed 8D Gaussian process does not take advantage of the spectrally banded features of the data that were pointed out in Section 3.1.1. The choice does have, however, a physical interpretation. Since the mean does not contain any information the difference in data can only be contained in the second-order statistics, namely the covariance. Looking at the broadband covariance matrix associated to an operating condition yields information about the *rms* vibrational power on each sensor (the diagonal terms of the covariance matrix) as well as the *rms* cross-power between sensors. Since the various sensors are distributed equally around the gearbox it can be concluded that this model provides information on proximity (i.e. relative power) and multipath effects (i.e. events that lead to high correlation between sensors). As shown by Bunks et al. (Bunks et al., 2000), this model results in an extremely robust classification of the Westland helicopter gearbox data.

To measure the ability of the 68 Gaussian models to discriminate the data they were used in three classification experiments. The experiments were conducted by testing which of the 68 models maximized the likelihood of observed data. This was done with varying sample lengths for the observed data (10k, 1k, and 200 samples). The three tests were performed using the fourth second of data. Note that since the models were created using the first 0.1 s of data that these tests are performed by applying the models to data removed by 3.9 s. The results of the above three experiments are illustrated in Fig. 1.

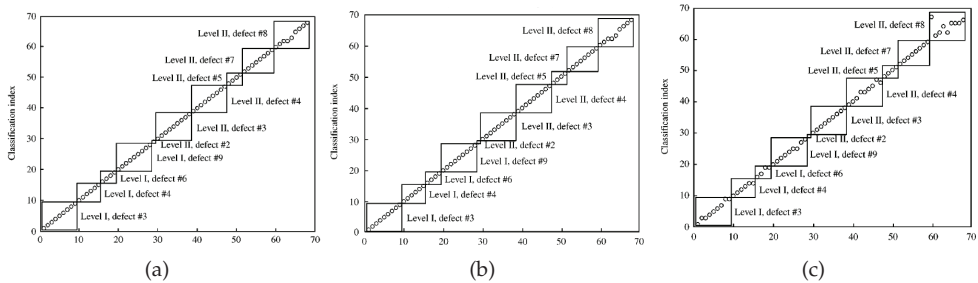


Fig. 1. Classification test using fourth second's first (a) 10k samples. (b) 1k samples. (c) 200 samples (Bunks et al., 2000).

Each sub-figure illustrates the classification index (1-68) vs the data set index (also 1-68). Thus, each circle in Fig. 1 (a) yields the result of deciding which is the best model (classification) given the data set specified on the independent axis. Perfect classification requires that all the circles lie on the diagonal as they almost do in Fig. 1 (a). The classification indices were chosen to group operating conditions together by defect type. The ensemble of rectangular boxes in Fig. 1 (a) indicates groups of indices whose defect type and level are the same. As shown in the figure each rectangle is labelled by the defect type and level it represents. Finally, inside each rectangle the torque level is increasing from the lower left to the upper right of the box. As can be seen in Fig. 1 (a) the classification of the 68 cases only has minor errors consisting of two torque-level misclassifications. In particular, for the level 2 defect d8 the torque at

50% is misclassified to be 45% and that of 60% is taken to be 50%. Thus, it can be concluded that using eight-dimensional Gaussian distributions to model the data is sufficient to identify defect types and almost sufficient to identify all torque levels. This remains true when the observed data set is reduced to only 1000 samples as is seen in Fig. 1 (b). However, as shown in Fig. 1 (c), when the number of samples is reduced to only 200 the classification of the 68 cases starts to show a more significant number of errors. Nevertheless, these errors continue to consist of only torque-level classification errors and none are defect-type errors.

The next test shows how the above classification procedure might work on a data set from an operating helicopter gearbox. The tests are made using a composite data set which was created by splicing together data from 10 of the 68 different operating conditions. An example of this data set for the port quill shaft sensor is illustrated in Fig. 2.

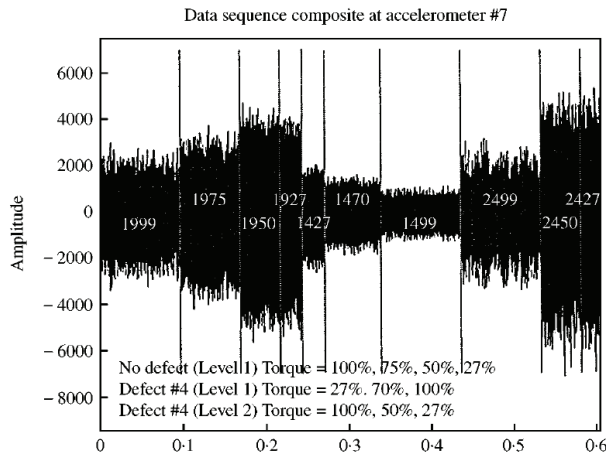


Fig. 2. Illustration of sequenced data from port quill shaft sensor.

The figure shows the 10 segments delineated by vertical lines. Also shown on each of the segments is a four-digit code of the form $XYZZ$ for which X gives the defect level (1 or 2), Y gives the defect type (1-9), and ZZ gives the torque level (the torque level designated by 99 actually indicates a torque of 100%). The sequence of segments is also summarized in the lower-left corner of the plot. The left-most segment of the data set is thus the no-defect case at 100% torque. This is followed by a diminishing of the torque to 75, 50, and then 27% at which point defect #4 (spiral bevel input opinion spalling) is introduced also at a torque level of 27%. The torque level is then increased to 70 and then 100% with this defect. The situation then becomes more serious evolving to a level-2 defect #4 at 100% torque. The torque level then decreases to 50 and then 27% for this defect level.

It is interesting to note that in the sequence presented in Fig. 2 the amplitude of vibration is higher for low torque levels than for high torque levels. This was already noted previously in the examination of power spectra (see Fig. 1 and the associated comments). The sequenced data set for all eight sensors (of which only one is illustrated in Fig. 2) were used to perform three classification experiments. Each experiment was conducted by applying a sliding window (with no overlap) of length 1k, 200, and 100 samples, respectively, over the data. The models used in the prior examples were used to classify the data output from the sliding window at each step. The results are shown in Fig. 3.

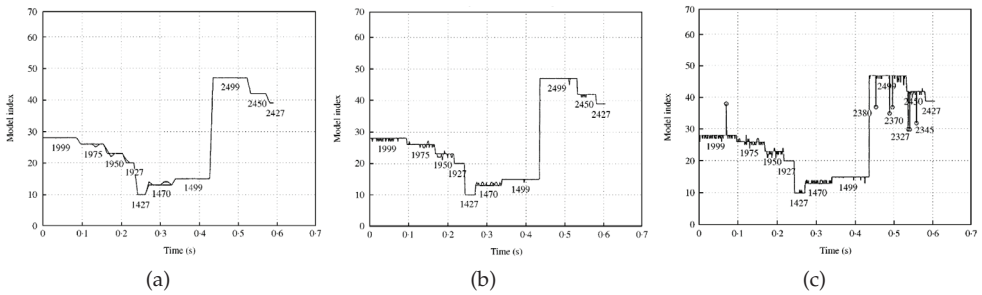


Fig. 3. Comparison of classification and true sequence with (a) a sliding window 1000 samples wide. (b) a sliding window 200 samples wide. (c) a sliding window 100 samples wide (Bunks et al., 2000).

In Fig. 3 (a) there are very few errors of classification for the case where a sliding window of 1k samples is used. Since model index is organized by torque as was illustrated in Fig. 1 it can be seen that the several small errors are torque-level classification errors only. For the case of the sliding window using 200 samples shown in Fig. 3 (b) there are more errors but they remain small and are still only slight torque-level discrepancies. Finally, in Fig. 3 (c) which illustrates the case of the sliding window with 100 sample points, there are several important errors which give rise to misclassifications in the type of defect. These errors are indicated by circles and are labelled by their type codes.

The tests presented in the preceding figures indicate that the 68 models are quite robust and are quite capable of performing accurate classification of defects and even classification of torque level. Furthermore, the examples show that the use of more samples is preferable since it improves the accuracy of classification and proportionally diminishes the amount of data to consider. It is important to note that the computational burden necessary to obtain the decision statistics is quite light depending only on a number of multiplies proportional to the number of samples as defined in Equation 2.

3.1.2.2 Modelling state transition probabilities

The construction of a state transition model improves classification performance over that obtained from simply using a finite set of unrelated models. That is to say the additional information obtained from knowing what are the relative state visitation frequencies provides a smoothing mechanism to the estimation of state (i.e. classification). Detection and estimation algorithms based on *HMMs* incorporate state transition information in a Bayesian way.

In the previous section, a description of how to model state probability densities for the Westland helicopter gearbox data was described and the resulting models were shown to be effective in discriminating between various types of seeded defects. The models were also shown to be robust and sufficiently discriminating to determine the operating torque levels of the gearbox. These results seem to be sufficiently good to forego the need for the construction of a state transition model. However, the Westland data set is a high-quality laboratory test set. Under true helicopter-operating conditions the recorded data would certainly be a lot noisier and would most likely contain features not apparent in the test data.

Another important point is that for general problems of machine health monitoring there will not always be laboratory test data from which state probability densities can be estimated as was done in the previous subsection. Under these circumstances it may be necessary

to determine these densities from explicit modelling using only the knowledge of the physical characteristics of the various machine components (for a description of how to do this with gears see (Mark, 1992)). Consequently, additional statistical information obtained from the state transition model is very important for the performance of any classification computations.

This section discusses a state transition model for the gearbox data. This model should be based on prior information about the frequency of occurrence of each defect and the average time spent operating at each of the torque levels. Although, for the Westland helicopter gearbox data set, this information is not available it is still useful to discuss a potential model. The model describes how a simple state transition diagram might reasonably be constructed. The model is illustrated in Fig. 4 which shows three rows of states the top one for no-defect, the middle one for a level 1 defect, and the bottom one for a level 2 defect.

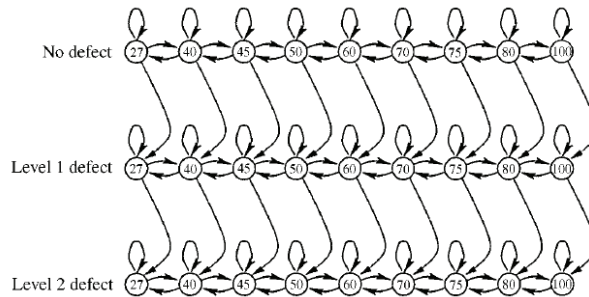


Fig. 4. Transition diagram for Westland helicopter gearbox.

Each row contains nine states representing the nine torque levels under which the Westland helicopter gearbox was operated. The arrows between torque levels are associated with probabilities for transitioning from one torque level to the next. As shown in the diagram they indicate that the torque can only increase or decrease by increments. The torque level can also stay the same. The arrows between defect levels are associated with probabilities for transitioning from no defect to a level-1 defect and subsequently to a level-2 defect. Notice that once a no-defect state is left there is no way to return. This is also true for the transition from a level-1 defect to a level-2 defect.

The state diagram in Fig. 4 illustrates a model for transitions from a no-defect to a single-defect case. Since the Westland data set consists of seven distinct defect types the full-state transition diagram would be based on Fig. 4 where the bottom two rows would be repeated six more times and each attached in parallel to the top row. The full-state transition model would then permit a transition from the no-defect case to any one of the seven defect cases. The state transition model could then be made more complex by allowing multiple defects to be simultaneously present. Under these circumstances, additional branches representing the multiple defect cases would be added to the state transition diagram. The new branches would be appropriately attached to the single defect states. For example, simultaneous defects A and B could be accessed by transitioning from either the single defect A or the single defect B (not necessarily with the same probability).

3.1.2.3 Prognostics

An important problem in CBM is planning essential maintenance. Maintenance must be performed before the expiration of the remaining useful life of critical machine components.

Nevertheless, performing machine maintenance before it is needed gives rise to unnecessary costs and can lead to maintenance-caused defects (damage inflicted on the machine via the maintenance procedure). Thus, the estimation of mean useful life, known as *prognostics*, is an interesting and important problem. *HMMs* provide a natural framework within which problems of *prognostics* can be formulated. To illustrate this, Fig. 5 shows an HMM which can be used for estimating the mean remaining useful life. In the figure, each state represents the degree of incipient failure of a particular machine defect. The random process of each state is determined by a row of Fig. 4. That is, state 1 in Fig. 5 is associated to a random process generated by the top row of Fig. 4 (i.e. for no-defects). State 2 in Fig. 4 gets its random process from the middle row of Fig. 4 and so on. For each state n there is a probability of remaining in that state, a_{nn} , and a probability of leaving it, $a_{n,n+1}$. Since defects always become worse and never improve the chain only allows transitions in the direction of increasing defect.

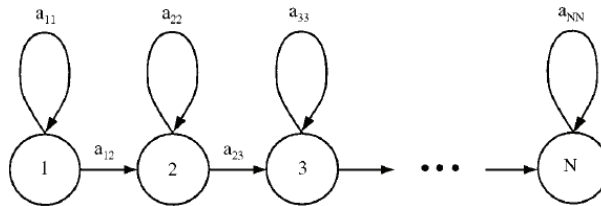


Fig. 5. Hidden Markov model for prognostics.

Thus, if state N represents the state of zero remaining useful life then the mean time, t^* , to this state given that the current state, n , is calculated as a function of the mean number of steps required to go from state n to state N . From the definition of expectation this is accomplished by calculating the probability, p_k , of going from state n to state N in exactly k steps and then forming the sum

$$t^* = \sum_{k=1}^{\infty} k p_k \quad (3)$$

This shows, at least in principle, how the problem of prognostics can be formulated if the appropriate transition probabilities can be determined. Transition probabilities are determined from experiential knowledge. As an example, it is well known that timing belts on cars fail at around 70 000 miles. This sort of information is known for many types of machine components and can be incorporated into the HMM in Fig. 5 as transition probabilities. Used in conjunction with observations of vibration data it helps to determine the current state in the HMM and the transition probabilities can then be used to compute a mean remaining useful life estimate. This also illustrates how *HMMs* can be used in a hierarchical way to build up complex models from simpler ones.

This application example does not answer all questions of how to apply *HMMs* to *CBM* and many issues remain to be resolved. In particular, it is important to explore the practical question of how to train *HMMs* for *CBM*. Since it is impractical to test each machine for each operating condition that might be encountered, Bunks et al. (Bunks et al., 2000) suggested a dynamic modelling approach for synthetically generating vibration data.

3.2 Sleep apnea diagnosis using HMMs

In this example, an automatic diagnosis approach based on *HMMs* is proposed for the diagnosis of *sleep apnea syndrome* (Al-ani et al., 2004; 2008).

Sleep apnea Syndrome (SAS) (American Academy of Sleep Medicine Task Force, 1999), is a very common *sleep disorder*. SAS is considered as clinically relevant when the breath stops during more than 10 seconds and occurs more than five times per sleep hour. These non breathing episodes may sometimes occur more than 300 times a night. Health studies affirm that more than 30 of these non breathing episodes per night should be considered abnormal. There exist two kinds of *apneic events* that may cause insufficient pulmonary ventilation during *sleep apnea* and *hypopnoea*. *Apnea* is defined as the total absence of airflow, followed by the reduction of oxygen levels in arterial blood. The term *hypopnoea* is used when the breath doesn't stop but decrease over 50% of its normal value, followed by the reduction of oxygen levels in arterial blood. The SAS is present mainly in adults and in 11% of children especially in the male population (Eliot et al., 2003). Different form of *apnea/hypopnoea* may be distinguished: *obstructive*, and *mixed*. *Obstructive sleep apnea (OSA)* is the most common type of *sleep apnea*. OSA occurs when the upper airway occludes (either partially or fully) but the effort to continue breathing is persistent. The primary causes of upper airway obstruction are lack of muscle tone during sleep, excess tissue in the upper airway, and anatomic abnormalities in the upper airway and jaw.

The treatment of SAS depends on the diagnosis quality of the *apnea/hypopnoea* event by the expert physician. Nowadays the *sleep apneas* are classified manually by the expert physician thanks to the nocturnal *polysomnographic* monitoring that simultaneously records several vital signals during the entire sleeping process (*Nasal airflow (NAF)*, *electrocardiogram (ECG)*, *electroencephalogram (EEG)*, *electromyogram (EMG)*, *esophageal pressure (Pes)*, *gastric pressure (Pgas)*, *Oxygen Saturation (OS)*, ...). A *sleep apnea diagnosis* is a very time consuming, expensive and tedious task consisting of expert visual evaluation all 10 minutes pieces of approximately 8 hour recording with a setting of many channels.

Al-ani et al. (Al-ani et al., 2004), demonstrated that *sleep apnea* classification may be done automatically using three simultaneous records of *NAF*, *Peso* and *Pgas* issued from the current techniques of investigating patients with suspected sleep disordered breathing. The inference method of this approach translates parameter values into interpretations of physiological and pathophysiological states. If the interpretation is extended to sequences of states in time, a state-space trajectory may be obtained. The identification of state-space trajectories is a useful concept in diagnosis because some disorders may only be distinguished from each other by time sequences of pathophysiological states. The probabilistic concept of *HMMs* captures the uncertainty inherent in state interpretation. A major interest of using *HMMs* is their capabilities for predictive inference. This inference makes our diagnosis system useful in the evaluation of treatment plans, in the optimization of treatment plans, in the predictive alarming.

The proposed diagnosis system, Fig. 6, is organized in such a manner that to be interactive with the expert. It is organized as three phases: training phase, state interpretation phase and detection phase.

3.2.1 Training phase

In the training phase, a real polysomnographic clinical records (observations) were used. These records were organised as a *multiple observation sequences* of observations vectors. Each vector contains three vital signals: *NAF*, *Peso* and *Pgas*. In Patient monitoring, we may represent the evolution of the processes, modelled by different *HMMs* (*normal respiration*, *snoring*, *hypopnea*, *apnea*, ...), by some intrinsic hidden states corresponding to some events of interest (e.g., two events *normal inspiration* and *normal expiration* in the case of normal

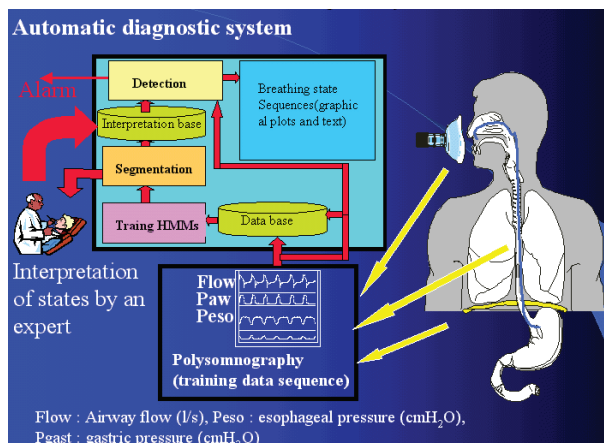


Fig. 6. Diagnosis system.

respiration. *Simulated annealing algorithm* (Al-ani & Hamam, 2010b) was used to construct these *HMMs*.

3.2.2 Segmentation of the training data sequence into different pathophysiological state sequence - state interpretation phase

Given the training observation sequences and their corresponding *HMMs*, this phase allows an expert to decode, off-line, their corresponding state labels using *Viterbi decoding algorithm*. The expert may then interpret these decoded state labels by some *pathophysiological* state interpretations (e.g. snoring, inspiration, expiration, obstructive apnea, etc.), Fig. 7. The interpretation of each state will be useful in on-line detection and interpretation of the on-line state sequence. The identification of state-space trajectories is useful concept in diagnosis since some disorders can only be distinguished from each other by the sequences of pathophysiological states that they follow in time.

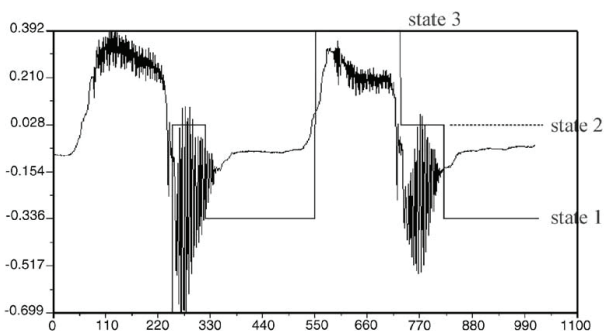


Fig. 7. An example of data segmentation of the air flow into different states and their interpretations. Event model contains three states: state 1: week airflow, state 2: snoring during expiration and state 3: snoring during inspiration.

3.2.3 Off-line or on-line detection phase

Once the *HMMs* are available, the detection of hidden *pathophysiological* states in a given observation sequence becomes possible. This may be done by two consecutive steps. At step 1, select the model that gives the maximum likelihood and at step 2, detect the *Pathophysiological* state using *Viterbi algorithm*. Two modes of detection are possible: off-line detection using the complete length of the observation sequence or on-line detection using a sliding window on the observation sequence stream.

In the above work, it was demonstrated that *sleep apnea* classification may be done automatically using three simultaneous records of *NAF*, *Peso* and *Pgas* issued from the current techniques of investigating patients with suspected sleep disordered breathing. Correctly identifying obstructive hypopneas and episodes of upper airway resistance needs a sensitive measure of airflow and inspiratory effort. The measurement of swings in pleural pressure by esophageal manometry is the current gold standard techniques for detecting changes in respiratory effort. However, the placement of an esophageal catheter is often uncomfortable and unacceptable, it may modify the upper airway dynamics, and some believe that it contributes to the sleep disturbance during the sleep study. Furthermore, this technique is available in only a proportion of sleep laboratories and, if performed, adds significantly to the cost of the sleep study. For all these reasons, a new preliminary approach on detecting and classifying *sleep apneas* and other *breathing disorders* is realised using mainly the *ECG* (Al-ani et al., 2008). In this work, the observation sequences were obtained by feature extraction technique which is based on *Heart Rate Variability (HRV)* defined in term of *RR intervals* of *ECG* signals.

4. Conclusions

We tried throughout this chapter to sensitize the reader

1. on two problems related to conventional *HMMs*: the training problem and the selection of a convenient *structure* for the constructed *HMM*. We recommend the reader to explore some of the new training schemes introduced in section 2.2.3 that may significantly improve the modelling results. The hybrid *generative/discriminative* approach is more sensitive to change detection in dynamic systems than the purely *generative* or the purely *discriminative* methods.

Concerning the choice of the *structure*, when it have a clear domain interpretation, as for example in fault monitoring, the *structure* may be naturally dictated by the domain. When the *structure* is not so conveniently available, however, we can employ some of the data-driven selection methods introduced in section 2.3 to discover an appropriate *structure*.

2. to the use of *HMMs* as a powerful tool for dynamic system modelling, fault detection and diagnosis applications.

5. References

- Aarts, E. & Korst, J. (1990). *Simulated Annealing and Boltzmann Machines - A Stochastic Approach to Combinatorial Optimization and Neural Computing*. J. Wiley & Sons, New York.
- Al-ani, T.; Hamam, Y.; Fodil, R.; Lofaso, F. & Isabey, D. (2004). Using hidden Markov models for sleep disordered breathing identification. *Simulation Practice and Theory*, Vol. 12, No. 2, pp. 117-128.

- Al-ani T. & Hamam, Y. (2006). A learning based stochastic approach for fault diagnosis in a continuous stirred tank reactor. MESM'06, Alexandria, August 28-30.
- Al-ani, T.; Karmakar, C.K. ; Khandoker, A.H. ; Palaniswami, M. (2008). Automatic Recognition of Obstructive Sleep apnoea Syndrome Using Power Spectral Analysis of Electrocardiogram and Hidden Markov Models. International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Sydney, NSW, 15-18 Dec., pp. 285-290.
- Al-ani, T. & Trad, D. (2010). Signal Processing and Classification Approaches for Brain-computer Interface", In : Intelligent and Biosensors, Edited by Vernon S. Somerset, pp. 25-66.
- Al-ani, T. & Hamam, Y. (2010). A low Complexity Simulated Annealing Approach For Training Hidden Markov Models. International Journal of Operational Research, Vol. 8, No.4 pp. 483-510.
- Altun, Y.; Tsochantaridis, I. & Hofmann, T. (2003). Hidden Markov support vector machines. In: Proceedings of the 20th International Conference on Machine Learning (ICML-2003), Washington, DC.
- American Academy of Sleep Medicine Task Force. (1999). Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The report of an American Academy of Sleep Medicine Task Force. Sleep, Vol. 22, No. 5, pp. 662-689.
- Andrieu, C. & Doucet, A. (2000). Simulated annealing for maximum a posteriori parameter estimation of hidden Markov models. IEEE Transactions on Information Theory, Vol. 46, No. 3, pp. 994-1004.
- Arslan, L.M. & Hansen, J.H.L. (1999). Selective Training in Hidden Markov Model Recognition. IEEE Transactions on Speech & Audio Processing, Vol. 7, No. 1, pp. 46-54.
- Aupetit, S.; Monmarche, N. & Slimane, M. (2007). Hidden Markov models training by a particle swarm optimization algorithm. Journal of Mathematical Modelling and Algorithms, Vol. 6, No. 2, pp. 175-193.
- Bach, F.R. & Jordan, M.I. (2005). Discriminative Training of Hidden Markov models for Multiple Pitch Tracking. In ICASSP.
- Bahl, L.R., Brown, P.F., De Souza, P.V. & Mercer, R.L. (1986). Maximum mutual information estimation of hidden Markov model parameters for speech recognition. In: Proceedings of IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP'86), Tokyo, Japan, pp. 49-52.
- Baldi, P. & Brunak, S. (2001). Bioinformatics. A Bradford Book, The MIT Press, Cambridge.
- Baum, L.E. & Eagon, J.A. (1967). An inequality with application to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. Accounting, Organizations and Society, Vol. 21, Issue 2, pp. 360-363.
- 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, Vol. 41, No. 1, pp. 164-171.
- Bunks, C.; McCarthy, D. & Al-ani, T. (2000). Condition-based maintenance of machines using hidden Markov models. Mechanical Systems and Signal Processing, Vol. 14, No. 4, pp. 597-612.

- Bicego, M.; Murino, V. & Figueiredo, M.A.T. (2003). A sequential pruning strategy for the selection of the number of states in Hidden Markov Models. *Pattern Recognition Letters*, Vol. 24, pp. 1395-1407.
- Biem, A.; Jin-Young, Ha & Subrahmonia, J. (2002). A Bayesian model selection criterion for HMM topology optimization. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Orlando, FL, USA. Proceedings, Vol. 1 pp. I-989 - I-992.
- Biem, A. (2003). A Model Selection Criterion for Classification: Application to HMM Topology Optimization. *Seventh International Conference on Document Analysis and Recognition*. Proceedings, Vol. 1, pp. 104-108.
- Castellani, A.; Botturi, D.; Bicego, M. & Fiorini P. (2004). Hybrid HMM/SVM :Model for the Analysis and Segmentation of Teleoperation Tasks. *Proceedings of the 2004 IEEE International Conference on Robotics & Automation New Orleans, LA, April 2004*.
- Celeux, G. & Durand, J.-B. (2008). Selecting hidden Markov model state number with cross-validated likelihood. *Computational Statistics*, Vol. 23, No. 4, pp. 541-564.
- Cheng, C.-C.; Sha, F. & Saul, L.K. (2009). A fast online algorithm for large margin training of continuous-density hidden Markov models. In *Proceedings of the Tenth Annual Conference of the International Speech Communication Association (Interspeech-09)*, pp. 668-671. Brighton, UK.
- Cheng, C.-C.; Sha, F. & Saul, L.K. (2010). Online Learning and Acoustic Feature Adaptation in Large Margin Hidden Markov Models. To appear in *IEEE Journal of Selected Topics in Signal Processing*.
- Daidone, A.; Di Giandomenico, F.; Chiaradonna, S. & Bondavalli, A. Hidden Markov Models as a Support for Diagnosis: Formalization of the Problem and Synthesis of the Solution. *25th IEEE Symposium on Reliable Distributed Systems, SRDS '06*.
- Do, T-M-T & Artières, T. (2009). Large Margin Training for Hidden Markov Models with Partially Observed States. *Proceedings of the 26th International Conference on Machine Learning, Montreal, Canada, 2009*.
- Dong, Y. & Li, D. (2007). Large-Margin Discriminative Training of Hidden Markov Models for Speech Recognition. *Proceedings of the International Conference on Semantic Computing*, pp. 429-438.
- Dunmur, A.P. & Titterton, D.M. (1998). The influence of initial conditions on maximum likelihood estimation of the parameters of a binary hidden Markov model. *Statistics & Probability Letters*, Vol. 40, No. 1, pp. 67-73.
- Eddy, S.R. (1995). Multiple alignment using hidden Markov models. *Proc. Third International Conference on Intelligent Systems for Molecular Biology*, C. Rawlings et al., eds. AAAI Press, Menlo Park, Vol. 3, pp. 114-120.
- Eliot, S.K.; Janita L.; Cheryl B. & Carole L.M. (2003). Transit Time as a measure of arousal and respiratory effort in children with sleep-disorder breathing. In: *Pediatric research*, Vol. 53, No. 4, pp. 580-588.
- Elliot, R. J.; Aggoun, L.; & Moore, J. B. (2004). Hidden Markov Models: Estimation and Control, Vol. 29 of *Applications of Mathematics*. New York: *Springer-Verlag*.
- Fengqin, Y. & Changhai, Z. (2008). An effective hybrid optimization algorithm for HMM. *Proc. of the 2008 Fourth International Conference on Natural Computation*, Vol. 4, pp. 80-84.
- Fink, G. A. (1989). *Markov Models for Pattern Recognition. From Theory to Applications*, Springer.

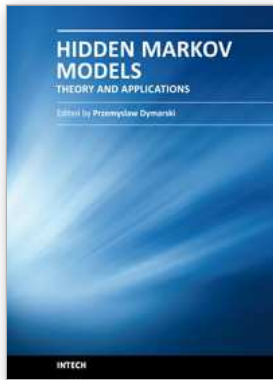
- Frankel, J. Linear Dynamic Models for Automatic Speech Recognition. Ph.D. Thesis, The Centre for Speech Technology Research, University of Edinburgh, Edinburgh, UK.
- Fraser, A.M. Hidden Markov Models and Dynamical Systems, Cambridge University Press.
- Ganapathiraju, A.; Hamaker, J. E. & Picone, J. (2000). Hybrid SVM/HMM architectures for speech recognition. In Proc. ICSLP-2000, Vol. 4, pp. 504-507, Beijing, China.
- Golowich, S.E. & Sun, D.X. (2007). A Support Vector/Hidden Markov Model Approach to Phoneme Recognition. Applied Soft Computing. Vol. 7, No. 3, pp. 828-839.
- Gu, L.; Nayak, J. & Rose, K. (2000). Discriminative training of tied-mixture HMM by deterministic annealing. ICSLP 2000, Beijing, China, October.
- Gunter, S. & Bunke, H. (2003). Optimizing the Number of States, Training Iterations and Gaussians in an HMM-based Handwritten Word Recognizer. Seventh International Conference on Document Analysis and Recognition, 3-6 Aug. Proceedings, Vol. 1, pp. 472-476.
- Heck, L.P. & McClellan, J.H. (1991). Mechanical system monitoring using hidden Markov models. Proceedings International Conference on Acoustics, Speech and Signal Processing, Toronto, Ont., Canada., Vol. 3, pp. 1697-1700.
- Helmy, S.; Al-ani, T.; Hamam, Y. & El-madbouly, E. (2008). P300 Based Brain-Computer Interface Using Hidden Markov Models. International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Sydney, NSW, 15-18 Dec., pp. 127-132.
- Hirsch H-G (2001). HMM adaptation for applications in telecommunication. Speech Communication, Vol. 34, Issues 1-2, April 2001, pp. 127-139.
- Hofbauer, M.W. Hybrid Estimation of Complex Systems. Lecture Notes in Control and Information Sciences, Vol. 319. Thoma, M. & Morari, M. (Editors).
- Huda, Md.S.; Ghosh R. & Yearwood J. (2006). A Variable Initialization Approach to the EM Algorithm for Better Estimation of the Parameters of Hidden Markov Model Based Acoustic Modeling of Speech Signals. Lecture Notes in Artificial Intelligence. Advances in Data Mining: Applications in Medicine, Web Mining, Marketing, Image and Signal Mining, Vol. 4065, pp. 416-430, Springer-Verlag, Berlin.
- Jaakkola, T. & Haussler, D. (1998). Exploiting generative models in discriminative classifiers. In: Proceedings of Advances in Neural Information Processing Systems, vol. 11.
- Jaakkola, T.; Meila, M. & Jebara, T. (1999). Maximum entropy discrimination. In: Proceedings of Advances in Neural Information Processing Systems, vol. 12.
- Jebara, T., Pentland, A. (1998). Maximum conditional likelihood via bound maximization and the CEM algorithm. In: Proceedings of Advances in Neural Information Processing Systems, Vol. 11.
- Jiang, H.; Li, X. & Liu, C. (2006a). Large Margin Hidden Markov Models for Speech Recognition. IEEE Transactions On Audio, Speech, and Language Processing, Vol. 14, No. 5, pp. 1584-1595.
- Jordan, M.I. (2004). Graphical models. Statistical Science 19 (Special Issue on Bayesian Statistics), pp. 140-155.
- Jiang, H.; Li, X. & Liu, C. (2006b). Solving Large Margin Estimation of HMMS Via Semidefinite Programming. In: Proc. Interspeech, pp. 2414-2417.
- Jiang, H. (2010). Discriminative training of HMMs for automatic speech recognition: A survey. Computer Speech and Language, Vol. 24, pp. 589-608.

- Kinney, G.W.; Barnes, J.W. & Colletti, B.W. (2007). A Reactive Tabu Search Algorithm with Variable Clustering for the Unicost Set Covering Problem. *International Journal of Operational Research*, Vol. 2, No. 2, pp. 156-172.
- Kirkpartrick, S.; Gelatt, C.D. & Vecchi, Jr. M.P. (1983). Optimization by simulated annealing, *Science*, Vol. 220, No. 4598, pp. 671-680.
- Kanevsky, D. (2004). Extended Baum Transformations for General Functions. *Proceeding ICASSP 2004*, Vol. 1, pp. 821-4.
- Kennedy, J.; Eberhart, R.C. & Shi, Y. (2001). *Swarm intelligence*. Morgan Kaufmann Publishers Inc., San Francisco, CA.
- Koloydenko, A. (2007). On adjusted Viterbi training. *Acta Applicandae Mathematicae*, Vol. 9, No. 2, pp. 111-126.
- Krüger S.E.; Schafföner M.; Katz M.; Andelic E. & Wendemuth A. (2005). Speech recognition with support vector machines in a hybrid system. In *Proc. EuroSpeech*, pp. 993-996.
- Krüger S.E.; Schafföner M.; Katz M.; Andelic E. & Wendemuth A. (2006). Mixture of Support Vector Machines for HMM based Speech Recognition. In *18th International Conference on Pattern Recognition (ICPR)*.
- Kwon1, K.-C; J.-H. Kim2 & Seong, P.-H. (2002). Hidden Markov model-based real-time transient identifications in nuclear power plants. *Special Issue: Intelligent Systems for Process Monitoring and Diagnosis*, Vol. 17, No. 8, pp. 791-811.
- Lee, J.-S. & Park, C.H. (2007). Training hidden Markov models by hybrid simulated annealing for visual speech recognition. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, Vol. 1, pp. 198-20.
- Lee, J-S & Park, C.H. (2005). Discriminative Training of Hidden Markov Models by Multiobjective Optimization for Visual Speech Recognition. *Proceedings of International Joint Conference on Neural Networks*, Montreal, Canada, July 31 - August 4, pp. 2053-2058.
- Lee, Z.-H. (2008). On-line MTO lead-time scheduling: a probabilistic approach. *The International Journal of Operational Research*, Vol. 3, No. 1-2, pp. 183 - 200.
- Li, L. & Noble, W.S. (2003). Combining Pairwise Sequence Similarity and Support Vector Machines for Detecting Remote Protein Evolutionary and Structural Relationships. *Journal of computational biology*, Vol. 10, No. 6.
- Li, X.; Jiang, H. (2005). A Constrained Joint Optimization Method for Large Margin HMM Estimation. *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, Cancun, Mexico, November 27 - December 1.
- Li, X.; Jiang, H. & Liu C. (2005a). Large Margin HMMS for Speech Recognition. *the 30th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, March 18-23.
- Li, X.; Jiang, H. & Liu C. (2005b). Discriminative Training of CDHMMs for Maximum Relative Separation Margin, *the 30th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, March 18-23.
- Li, Q. & Juang, B-H. (2006). Study of a Fast Discriminative Training Algorithm for Pattern Recognition. *IEEE Transactions on Neural Networks*, Vol. 17, No. 5, pp. 1212-1221.
- Liporace, L. A. (1982). Maximum likelihood estimation for multivariate observations of Markov sources. *IEEE Transactions on Information Theory IT-28*, pp. 729-734.
- Liu, J.; Wang, Z. & Xiao, X. (2007a). A hybrid SVM/DDBHMM decision fusion modeling for robust continuous digital speech recognition. *Pattern Recognition Letters*, Vol. 28, pp. 912-920.

- Liu, C.; Jiang, H. & Li, X. (2007b). Discriminative training of CDHMMS for maximum relative separation margin. *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 8, pp. 2383-2392.
- Lyu, J.J.; Gunasekaran, A. & Ding, J.H. (1996). Simulated Annealing Algorithms for Solving the Single Machine Early/Tardy Problem. *International Journal of Systems Science*, Vol. 27, No. 7, pp. 605-610.
- Paul, D.B. (1985). Training of hmm recognizers by simulated annealing. *Proceedings IEEE International Conference Acoustics, Speech and Signal Process*, Vol. 10, pp. 13-16.
- Pérez, O'; Piccardi, M.; Garcí'a, J.; Patricio, M.A. & Molina, J.M. (2007). Comparison between genetic algorithms and the Baum-Welch algorithm in learning HMMs for human activity classification. *Lecture Notes in Computer Science*, Vol. 4448, pp. 399-406, Springer Berlin.
- Poritz, A.B. (1982). Linear predictive hidden Markov models and the speech signal. *Proceedings of ICASSP, Paris, France, May*, pp. 1291-1294.
- Povey, D. & Kingsbury, B. (2007). Evaluation of Proposed Modifications to MPE for Large Scale Discriminative Training. *ICASSP'07*, Vol. 4, pp. 321-324.
- Mark, W.D. (1992). In *Noise and vibration Control Engineering: Principles and Applications*, pp. 735-770. New York: John Wiley & Sons. Elements of gear noise prediction.
- Markov, K.; Nakagawa, S. & Nakamura, S. (2001). Discriminative training of HMM using maximum normalized likelihood algorithm. In *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 497-500.
- Mazumdar, C.S., Mathirajan M., Gopinath, R. and Sivakumar A.I. (2007). Tabu Search methods for scheduling a burn-in oven with non-identical job sizes and secondary resource constraints. *International Journal of Operational Research*, Vol. 3, No. 1-2, pp. 119-139.
- McKendall Jr, A. R. & Shang, J. (2008). Tabu Search Heuristics for the Crane Sequencing Problem. *International Journal of Operational Research*, Vol. 3, No.4, pp. 412-429.
- Meir, R. (1995). Empirical risk minimization versus maximum likelihood estimation: a case study. *Neural Computation* Vol. 7, No. 1, pp. 144-157.
- Miao, Q.; Huang, H.-Z. & Fan, X. (2007). A Comparison Study of Support Vector Machines and Hidden Markov Models in Machinery Condition Monitoring. *Journal of Mechanical Science and Technology*, Vol. 21, pp. 607-615.
- Myers, C.; Singer, A.; Shin, F. & Church, E. (1992). Modeling chaotic systems with hidden Markov models. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol.4, 23-26 Mar, San Francisco, CA , USA, pp. 565-568.
- Novák, D.; Cuesta-Frau, D.; Al ani, T.; Aboy, M.; Mico, R. & Lhotska, L. (2004). Speech Recognition Methods Applied to Biomedical Signals Processing. *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBECE)*, San Francisco, CA, 1-5 Sept, pp. 118-121 , Vol.1.
- Novák, D.; Al-ani T.; Hamam, Y. & L. Lhotska, (2004). "Electroencephalogram processing using Hidden Markov Models", *5th EUROSIM Congress on Modelling and Simulation*, September, 06-10, Paris.
- Novák, D.; Al-ani T.; Hamam, Y.; Cuesta-Frau, D.; Aboy, M. & Lhotska, L. (2004). Morphology Analysis of Biological Signals Using Hidden Markov Models. *17th conference of the International Association for Pattern Recognition (ICPR)*, 23-26 August, Cambridge, UK.

- Novák, D.; Al-ani T.; Hamam, Y.; Cuesta-Frau, D.; Mico, P. & Lhotska, L. (2004). Unsupervised Learning of Holter ECG signals using HMM optimized by simulated annealing. In *Analysis of Biomedical Signals and Images*. Brno: VUTUM Press, pp. 60-62.
- Vaseghi, S. V. (2006). *Advanced Digital Signal Processing and Noise Reduction*. 4th Edition, Wiley.
- Rabiner, L.R.; Wilpon, J.G. & Juang B.H. (1986). A segmental K-mean training procedure for connected word recognition. *AT&T Technical J.*, Vol. 65, pp. 21-31.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, Vol. 77, pp. 257-286.
- Rao, A.V. & Rose, K. (2001). Deterministically annealed design of hidden Markov model speech recognizers. *IEEE Transactions on Speech and Audio Processing*, Vol. 9, No. 2, pp. 111-126.
- Rao, S. & Hong, J. (2010). *Analysis of Hidden Markov Models and Support Vector Machines in Financial Applications*. Technical Report No. UCB/EECS-2010-63.
- Rodríguez, L.J. & Torres, I. (2003). Comparative Study of the Baum-Welch and Viterbi Training Algorithms Applied to Read and Spontaneous Speech Recognition. *Lecture Notes in Computer Science, Pattern Recognition and Image Analysis*, Vol. 2652, pp. 847-857, Springer.
- Sanguansat, P.; Asdornwiset, W. & Jitapunkul, S. (2004). Online Thai handwritten character recognition using hidden Markov models and support vector machines. *IEEE International Symposium on Communications and Information Technology, ISCIT*.
- Saon, G. & Povey, D. (2008). Penalty function maximization for large margin HMM training. In *Proc. Interspeech*.
- Scholkopf, B. & Smola, A.J. (2002). *Learning with Kernels: Support vector Machine, Regularization, Optimization, and Beyond*. The MIT Press, Cambridge.
- Seymore, K.; Mccallum, A. & Rosenfeld, R. (1999). Learning Hidden Markov Model Structure for Information Extraction. In *AAAI 99 Workshop on Machine Learning for Information Extraction*, pp. 37-42.
- Sha, F. & Saul, L. K. (2009). Large Margin Training of Continuous Density Hidden Markov Models, in *Automatic Speech and Speaker Recognition - Large Margin and Kernel Methods*, Editor: Joseph Keshet, Samy Bengio.
- Siddiqi, S.M.; Gordon, G. & A. Moore. (2007). Fast State Discovery for HMM Model Selection and Learning, *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics. (AI-STATS)*, 2007.
- Sloin, A. & Burshtein, D. (2008). Support Vector Machine Training for Improved Hidden Markov Modeling. *IEEE Transactions On Signal Processing*, Vol. 56, No. 1, pp. 172-188.
- Smola, A.J.; Bartlett, P.; Scholkopf, B. & Schuurmans, D., (Ed.) (2000). *Advances in Large Margin Classifiers*, The MIT Press.
- Smyth, P. (1994). Hidden Markov models for fault detection in dynamic systems. *Pattern Recognition*, Vol. 27, pp. 149-164.
- Stolcke, A. & Omohundro, S. (1993). Hidden Markov model induction by Bayesian model merging. In S. J. Hanson, J. D. Cowan, and C. L. Giles, editors, *Advances in Neural Information Processing Systems 5*, San Francisco, CA. Morgan Kaufmann, pp. 11-18.
- Surendran, D. & Levow, G. (2006). Dialog Act Tagging with Support Vector Machines and Hidden Markov Models. *Proceedings of Interspeech 2006*, pp. 1950-1953.

- Takara, T.; Iha, Y. & Nagayama, I. (1998) SSelection of the Optimal Structure of the Continuous HMM Using the Genetic Algorithm. 5th Int. Conf. On Spoken Language Processing. 3, pp. 751-754.
- Tang, J.; Zhu, J.-H. & Sun, Z.-Q. (2006). Evolutionary hidden Markov modeling for dynamic agent systems. *Kongzhi yu Juece-Control and Decision*, Vol. 21, No. 2, pp. 189-192.
- Taskar, B.; Guestrin, C. & Koller, D. (2003). Max-margin Markov networks. In: *Proceedings of Neural Information Processing Systems Conference (NIPS03)*, Vancouver, Canada.
- Tsontzos, G.; Diakouloukas, V.; Koniaris, C. & Digalakis, V. (2007). Estimation of General Identifiable Linear Dynamic Models with an Application in Speech Recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.
- Turner, R. (2008). Direct maximization of the likelihood of a hidden Markov model. *Computational Statistics and Data Analysis*, Vol. 52, No. 9, pp. 4147-4160.
- Valstar, M.F. & Pantic, M. (2008). Combined Support Vector Machines and Hidden Markov Models for Modeling Facial Action Temporal Dynamics. *Proceedings of Belgium-Netherlands Conf. ArtificialIntelligence (BNAIC'08)*, pp. 367-368, Boekelo, the Netherlands, October.
- Van Trees, H. L. (1968). *Detection, Estimation, and Modulation theory: Part I*. New York: John Wiley & Sons.
- Viterbi, A.J. (1967). Error bounds for convolutional codes and asymptotically optimum decoding algorithm. *IEEE Transactions On Information Theory*, Vol. 13, No. 2, pp. 260-269.
- Xue, L.; Yin, J.; Ji, Z. & Jiang, L. (2006). A particle swarm optimization for hidden Markov model training. *8th International Conference on Signal Processing*, Vol. 1, pp. 16-20.
- Yang, F.; Zhang, C. & Bai, G. (2008). A novel genetic algorithm based on tabu search for HMM optimization. *4th International Conference on Natural Computation, ICNC '08*, 18-20 Oct., Vol. 4, pp. 57-61.
- Vapnik, V.N. (1998). *Statistical Learning Theory*. Wiley.
- Wang, X.-M.; Wang, Q. & Yao, T.-R. (2008). Training algorithm of hidden Markov model based on mixture of factor analysis. *Journal of System Simulation*, Vol. 20 , No. 15, pp. 3969-3972.
- Wren, C.R.; Clarkson, B.P. & Pentland, A.P. (2000). Understanding purposeful human motion. *Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000*. Proceedings. pp. 378-383
- Woodland, P.C. & Povey, D. (2002). Large Scale Discriminative Training of hidden Markov models for Speech Recognition. *Computer Speech and Language*, Vol. 16, No. 1, pp. 25-47.
- Won, K.-J.; Hamelryck, T.; Prügél-Bennett, A. & Krogh, A. (2007). An evolutionary method for learning HMM structure: Prediction of protein secondary structure. *BMC Bioinformatics*, Vol. 8, No. 357, doi:10.1186/1471-2105-8-357.
- Xu, W.; Wu, J.; Huang, Z. & Guan, C., (2005). Kernel based hidden Markov model with applications to EEG signal classification. *IASTED ICBE*, pp. 401-404.
- Xu, W.; Wu, J. & Huang, Z. (2006). A maximum margin discriminative learning algorithm for temporal signals. *The 18th International Conference on Pattern Recognition (ICPR'06)*
- Zhao, Z.-J; Zheng, S-L; Xu, C-Y & Kong, X-Z (2007). Discrete channel modelling based on genetic algorithm and simulated annealing for training hidden Markov model. *Chinese Physics*, Vol. 16, No. 6, pp. 1619-1623.



Hidden Markov Models, Theory and Applications

Edited by Dr. Przemyslaw Dymarski

ISBN 978-953-307-208-1

Hard cover, 314 pages

Publisher InTech

Published online 19, April, 2011

Published in print edition April, 2011

Hidden Markov Models (HMMs), although known for decades, have made a big career nowadays and are still in state of development. This book presents theoretical issues and a variety of HMMs applications in speech recognition and synthesis, medicine, neurosciences, computational biology, bioinformatics, seismology, environment protection and engineering. I hope that the reader will find this book useful and helpful for their own research.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Tarik Al-ani (2011). Hidden Markov Models in Dynamic System Modelling and Diagnosis, Hidden Markov Models, Theory and Applications, Dr. Przemyslaw Dymarski (Ed.), ISBN: 978-953-307-208-1, InTech, Available from: <http://www.intechopen.com/books/hidden-markov-models-theory-and-applications/hidden-markov-models-in-dynamic-system-modelling-and-diagnosis>

INTECH

open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.