NEW TRENDS IN MARKOV MODELS AND RELATED LEARNING TO RESTORE DATA

Florence Forbes *Wojciech Pieczynski*

ABSTRACT

We present recent approaches that extend standard Markov models and increase their modelling power. These capabilities are illustrated in the cited published works and more recently in the contributions to the Special Session on Markov models of the IEEE International Workshop on Machine Learning for Signal Processing, 2009. However, the review is not exhaustive and major older works may be missing.

1. INTRODUCTION

Markov models combine local relations to build stochastic models that exhibit great complexity. Such stochastic models have found applications in areas as diverse as document image analysis [77], remote sensing [18, 74], medical imaging [39, 49, 66, 76, 70], genetics [11, 44, 75], epidemiology [38], economics [6], etc. However, these powerful and flexible techniques are still restricted by several generic sources of complexity in data that require methods beyond the commonly-understood tools. Often data exhibit complex dependence structures, having to do for example with repeated measurements on individual items, or spatial or temporal association, and so on. Sections 2.1 to 2.3 provide directions toward the modelling of such dependencies. Other sources of complexity are connected with the measurement process, such as having multiple measuring instruments or simulations generating high dimensional (Section 2.5) and heterogeneous data (Section 2.4) or such that data are dropped out or missing (Section 2.6).
In Markov modelling and more generally in graphical modelling, the logical structure of a joint probability distribution is represented in the form of a network or graph depicting the local relations among variables. The graph can have directed (Bayesian networks, Section 2.9) or undirected (Markov random fields (MRF)) edges between the nodes, which represent the individual variables. When parts of the variables are not observed or missing, we refer to these models as Hidden Markov Models (HMM). In standard models, hidden variables are either continuous or discrete. However, a number of important applications require more flexibility to allow *fuzzy* or *mixed* states (Section 2.7).
Then, choosing the probabilistic model that best accounts for the observations is an important first step for the quality of the subsequent statistical inference and analysis. In section 2.10, we mention a number of criteria for choosing the most relevant models in a Markovian framework.

Regarding inference and learning of such models, spatial dependencies and potential hidden variables may generate a large amount of computation. Typically for MRFs, estimation algorithms correspond to an energy minimization problem (possibly under non standard constraints as in [72]) which is NP-hard and usually performed through approximation. Most approaches are based either on simulation methods such as Markov Chain Monte Carlo (MCMC) techniques (*e.g.* [65]) or on deterministic variational methods (*e.g.* [41]). While variational methods provide fast and reasonable approximate estimates, simulation methods offer more consideration of important theoretical issues such as accuracy of the approximation and convergence of the algorithms but at a much higher computational cost. As effective algorithms which show good performance in practice, we briefly discuss the well known Expectation-Maximization (EM) algorithm and its variational variants (Section 3.1) and the Iterative Conditional Estimation (ICE) method that uses simulations (Section 3.3). We then mention graph cuts (Section 3.2) and some recent hybrid approaches (Section 3.4) attempting to combine the main features and advantages of both simulation and deterministic methods.

2. RECENT TRENDS IN MARKOV MODELS

2.1. Pairwise and triplet Markov models

Let X and Y be respectively the hidden and observed processes. In standard HMMs, the distributions $p(x)$, $p(x, y)$, $p(x|y)$ are Markovian, whereas $p(y|x)$ is the distribution of independent variables. In Pairwise Markov Models (PMM), first introduced in the MRF context [61], the distributions $p(x, y)$, $p(y|x)$ and $p(x|y)$ are Markovian, whereas $p(x)$ is not necessarily Markovian. Thus $p(y|x)$ is more sophisticated in PMM and, as the Markovianity of $p(x|y)$ allows us to perform the same Bayesian processing, PMM are, in theory, more interesting. They are also in practice; as shown in [28], in the case of Markov chains the use of PMMs can significantly improve the results obtained with HMMs. Let

F. Forbes is with INRIA Grenoble Rhône-Alpes, Lab. Jean Kuntzman, Team Mistis, Grenoble, France (e-mail: Florence.Forbes@inrialpes.fr)

W. Pieczynski is with Telecom Sud Paris, Dept CITI, CNRS UMR 5157, Evry, France (e-mail: wojciech.pieczynski@it-sudparis.eu)

us notice that any HMM can, in principle, be extended to a PMM by transposing the Markovian structure of $p(x)$ to the Markovian structure of $p(x, y)$. PMMs can be extended to triplet Markov models (TMMs). Let us consider the case of triplet Markov chains (TMCs), where there are three random chains $X = (X_1, \ldots, X_N)$, $U = (U_1, \ldots, U_N)$ and $Y = (Y_1, \ldots, Y_N)$, with Y observed and X, U hidden. The chain X is the hidden signal one is looking for, and the chain U can have various interpretations, resulting in rich possibilities of modelling various problems [60]. The main interest of a TMC lies in the generality of its definition: only the triplet $T = (X, U, Y)$ is assumed to be Markovian, and none of the six chains $X, U, Y, (X, U), (X, Y), (U, Y)$ is necessarily Markovian. For discrete U and X, U can model the non stationarity of the chain (X, Y), the semi-Markovianity of X [4], or the evidential nature of the model [60, 57]. For continuous U and X, the classical Kalman filter can be extended to such Gaussian TMC [1]. The contribution [2] to the special session joins in this general problem. TMC can be extended to Triplet partially Markov chains (TPMC), in which the long memory noise [30] can be taken into account [46]. Let us also mention the triplet Markov fields (TMF) recently applied in textured images segmentation [5, 10].

2.2. Conditional random fields

There exists a relationship between Triplet Models and Conditional Random Fields (CRF) [45] which have been widely and successfully used in a number of applications (*eg.* [43]). CRFs are discriminative models in the sense that they model directly the conditional distribution of the labels given the observations. Explicit models of the joint distribution of the labels and observations or of the noise distribution are not required. In classification issues, the conditional distribution is the one needed. However, even in classification contexts, approaches that model the joint distribution are considered. They are known as generative models. Triplet Markov models belong to this class. Such generative models are certainly more demanding in term of modelling but they have the advantage to provide a model of the observed data (the likelihood) allowing this way better access to theoretical properties of the estimators.

2.3. Copulas and correlated non Gaussian noise

One of the limitations of the classical HMM (X, Y) open to criticism, is the simplicity of the "noise" distribution $p(y|x)$, in which the components of y are almost systematically considered to be independent. PMC and TMC are two successive extensions of $p(y|x)$. In PMC, $p(y|x)$ is Markovian and in TMC, $p(y|x)$ is the marginal distribution - not necessarily Markovian - of the Markovian distribution $p(u, y|x)$. Let us consider the PMC case and assume (X, Y) to be stationary: its distribution is then defined by $p(x_1, y_1, x_2, y_2) = p(x_1, x_2)p(y_1, y_2|x_1, x_2)$. To define each $p(y_1, y_2|x_1, x_2)$, we can use copulas which allow one to define a joint distribution from a pair formed by the marginal distributions and a dependence structure [55]. Let $h(y_1, y_2)$ be a probability density on R^2, H the associated cdf function, $h_1(y_1)$ and $h_2(y_2)$ the marginal densities, and H_1, H_2 the cdf functions associated with them. Then there exists a function C defined on $[0, 1]^2$, called a "copula", such that $H(y_1, y_2) = C(H_1(y_1), H_2(y_2))$. Conversely, having H_1, H_2 and a copula C, one can use the equation above to define H. This provides a very rich family of possible distributions for $p(y_1, y_2|x_1, x_2)$, each of which gives $p(y|x)$ [14]. The same can be said straigtforwardly for TMC replacing X by (X, U).

2.4. Theory of Evidence

The Dempster-Shafer "theory of evidence" (DSTE [73]) offers an interesting framework, which can be seen as an extension of the classical probability on finite space. It is especially interesting in the case of numerous sensors fusion, when they are of different nature. The use of the DSTE in the context of Markov models is quite recent and is undoubtedly very promising [35, 62]. A link with TMC [57] and TMF [59] has also been been established. Finally, "evidential filters" based on evidential measures have been proposed recently [29, 54]. Paper [63] deals with this issue.

2.5. Taking into account the curse of dimensionality

In HMMs, the multivariate Gaussian distribution is the most commonly used but it suffers from significant limitations when it comes to modelling real data sets. For very high dimensional data, the general covariance matrix model involves the estimation of too many parameters, leading to intractable computations or singularity issues. Solutions have been proposed based on so-called *parsimonious models* [20] but they are not specifically designed for high dimensional data. Other approaches consider reducing the dimensionality of the data as a pre-processing step possibly using Principal Component Analysis or variable selection methods. In a clustering context, this may not be satisfactory as relevant information may be lost that can help separating the classes. For these reasons, a more recent approach have been developed for independent Gaussian mixtures [12]. This approach has been extended to hidden MRFs in [10] where its efficiency and tractability has been illustrated on texture classification.

2.6. Missing observations

In practice, missing observations are very common, due for instance, to limitations of the measuring devices. The simplest solution in this case is to remove, from the data set, individuals for which observations are missing. This obviously results in weak analysis, in particular when a lot of

individuals are affected. A second very common approach is to replace the missing observations by some imputed values (*eg.* zeros, means) so as to carry out then a complete data analysis as usual. Although very popular, this technique tends to produce biased estimation. For independent mixtures, more sophisticated methods have been proposed (*eg.* [67] proposed to use the EM algorithm). More recently [11] proposed a framework to deal with missing observations within a MRF modelling.

2.7. Fuzzy and "mixed states" hidden Markov models

In general, the hidden chain X takes its values either from a continuous, or from a finite set. In the first case one generally considers that its distribution has some density with respect to the Lebesgue measure, which implies that the probability of any single point is null. In some situations, it is necessary to have simultaneously a "continuous" part, and a "discrete" part in this distribution. Then the distribution of X is given by a density with respect to a sum of the Lebesgue measure and a finite number of Dirac's measures. Such models were first introduced to model "fuzzy" classes and successfully used in the Markov fields context [49, 66, 68] and then in that of Markov chains [17, 18, 39, 47, 69]. More recently, it has been used in motion modelling and detection [23, 24].

2.8. Filtering and smoothing in switching linear systems

Let us consider a TMC with finite U and continuous X and Y. The problem of filtering in such a "switching" model consists of computing $p(u_n|y_1^n)$ and $E[X_n|u_n, y_1^n]$, when the problem of smoothing consists of computing $p(u_n|y_1^N)$ and $E[X_n|u_n, y_1^N]$. These quantities cannot be computed with complexity polynomial in time in the classical models, in which the distribution of $T = (X, U, Y)$ - which is a TMC - is given by a Markov distribution of (X, U) and by the conditional distribution $p(y|x, u)$. Thus different approximations, deterministic [22] or stochastic [3, 16, 64], have to be used. More recently, it has been shown that considering a TMC $T = (X, U, Y)$ in which (U, Y) is Markovian makes both computations feasible with complexity linear in time ([58] and references therein). Such models can be extended in numerous directions; in particular, copulas could be used in the PMC (U, Y).

2.9. Probabilistic Networks

Different Markov models (chains, fields) discussed above are particular cases of Probabilistic Networks (PNs) [8, 37, 40, 9]. Thus different extensions specified above can be studied in this more general framework with enormous possibilities of applications. The contribution [48] to the special session uses hidden Markov trees (HMT) and thus joins in this general problem. More generally, a link with multiresolution analysis can be made through Markov tree models. HMTs were originally proposed in [25] to detect changes in binary trees associated to wavelet coefficients for signal processing applications. The work in [31] has built on this by proposing a new upward-downward algorithm and a restoration algorithm to recover the hidden tree in a globally optimal way. Another set of Markov models belonging to the family of PNs is studied in [77].

2.10. Approximations for selecting models

When recasting the model choice as a probabilistic model comparison issue, most selection criteria involve calculating integrated (over the model parameters) likelihoods for a number of models. These integrated likelihoods are often high dimensional and intractable. Various approximations have been proposed. In particular the Bayesian Information Criterion (BIC) approximation of [71] is based on the Laplace method for integrals but many other approaches can be found in the literature (see the references in [42]). In the MRF context, most criteria are intractable. An approximated BIC based on the mean field approximation has been proposed in [33]. Other criterions exist but have not been applied to select Markov models (*eg.*DIC [19] and ICL [7]).

3. LEARNING IN MARKOV MODELS

Issues involving missing data are typical settings where exact inference is not tractable. We recall below three classes of algorithms commonly used in practice.

3.1. EM based method

The Expectation-Maximization (EM) algorithm is a general and now standard approach to maximization of the likelihood [27] in missing data problems. However, when focusing on HMRF estimation, difficulties arise due to the dependence structure in the model. An approach using mean field principle has been proposed in [79] and further developed in [21]. The mean field approach consists of using a simpler tractable model such as the family of independent distributions. Distributions from this family can be obtained by fixing the neighbors of each node to constant values. In [21], three different schemes for adaptively fixing the neighbors are proposed and compared.
The mean field approximation principle is among the variational methods [41] the simplest principle although not yet fully understood in some aspects. See also [52] for a summary of existing works. We observe that in many practical applications, the mean field approximation already handles a good part of the complexity of the data. There is a trade-off then between finer approximations which may not always lead to much better (classification or estimation) results and additional computational burden. Also, very few results exist on the quality of the variational approximations when they act as a surrogate in a larger inference problem.

In particular, as regards parameter estimation, results on bounds on the likelihood may not be of great help. Mean field methods can be related to *message passing* algorithms which correspond to general schemes for fitting variational approximations. Message passing algorithms include Mean Field, Loopy Belief Propagation, Expectation Propagation, Tree-reweighted message passing, Fractional Belief Propagation, Power Expectation Propagation. A unifying view and references for these different methods can be found in [51]. It appears that the difference between mean field methods and belief propagation methods is not the amount of structure they model but the measure of loss they minimize. Neither method is inherently superior. To our knowledge, there exists no experimental comparison of the various algorithms performance on real-world networks and data.

3.2. Graph Cuts

Apart from variational methods, recent efficient techniques such as *Graph cuts* [13] are built to provide Maximum A Posteriori solutions. However, they do not seem to provide the same flexibility as probabilistic techniques when parameter estimation is required. They are based on *hard membership* which allows the use of the graph cut optimization methods that have proven highly effective for solving labeling problems in some cases. EM approaches are based on *soft membership* adopting a more statistical point of view with some emphasize on parameter estimation and in particular on the biais inherent to hard membership methods. The shape of the likelihood function can be a major source of difficulties. An illustration and an attempt to study its properties in a HMM training framework is proposed in [78].

3.3. ICE method

Let us consider two random processes (V, Y) whose distribution depends on a parameter vector $\theta = (\theta_1, \ldots, \theta_m)$. ICE is an iterative method, to estimate θ from Y, based on the following principle. Let $\hat{\theta}(v, y)$ be an estimator of θ from complete data and let us assume that we can sample realizations of V according to $p(v|y)$. The ICE sequence is obtained as follows:
(i) Initialize θ^0;
(ii) Compute $\theta_i^{q+1} = E[\hat{\theta}_i(V, Y)|Y = y, \theta^q]$ for the components θ_i for which this is possible explicitly;
(iii) Simulate $v_i^q, \ldots, v_l^q$ according to $p(v|y, \theta^q)$ for the other θ_i and set: $\theta_i^{q+1} = [\hat{\theta}_i(v_1^q, y) + \ldots + \hat{\theta}_i(v_l^q, y)]/l$.
Thus ICE is applicable under two very weak hypothesis: (H1): the existence of an estimator $\hat{\theta}(v, y)$ from the complete data, and (H2): the ability to simulate V according to $p(v|y)$. As (H2) is verified for Markov models considered in this paper, the problem reduces to (H1). The idea behind ICE is very different from the idea leading to EM and is based on the following. Assuming that $\hat{\theta}(v, y)$ has a good small quadratic error, one wishes to approximate it by a function of the only observed variables y. The "best" - with regard to the same "quadratic error" criterion - approximation is the conditional expectation. As this expectation depends on the parameter, we arrive at the point (ii) in the definition of ICE above. Concerning the convergence of ICE, let us mention a recent theoretical result obtained in the case of independent data [56]. As in the case of EM, convergence can be obtained under some reasonable hypothesis if the initial value θ^0 is close enough to the real value θ. ICE was successfully used in numerous problems [5, 28, 46, 53, 59, 68], and the contribution [48] is also based on ICE. Let us notice that EM and ICE can give, in particular parametrizations of exponential models, the same sequence [26]. Besides, ICE and EM have been compared in the context of the classical HMCs and HMTs [53], and no noticeable difference has been observed. In conclusion, we may say that the performances of EM and ICE are comparable in studies performed until now, whereas ICE is easier to use.

3.4. Hybrid approaches

Variational methods have been shown to provide fast and reasonable approximate estimates in many scenarios [41] but frequently with little consideration of important theoretical issues. Convergence results exist for the Variational EM (VEM) algorithms [15] but their application is restricted to specific settings. Variants to extend the application domain of algorithms such as VEM have been proposed [79, 21] but they did not succeed in preserving the convergence results. Simulation methods appear then as natural candidates to make algorithms tractable for a wider class of problems while providing tools to study their convergence. As an example, the convergence of MCMC based algorithms has been widely studied and a lot of tools are now available that make various convergence results available or at least easy to derive (*eg.* [34] for MCEM). Recently [32] showed that combining both type of methods could greatly improve accuracy and modelling flexibility in missing data settings. As another attempt, in [36], a class of MCMC algorithms that use variational approximations as initial proposal distributions is introduced for sigmoidal belief networks.

4. REFERENCES

[1] B. Ait-el-Fquih and F. Desbouvries, Kalman filtering in triplet Markov chains, IEEE Trans. Sig. Proc., vol. 54, no 8, pp. 2957-2963, 2006.

[2] B. Ait-El-Fquih and F. Desbouvries, Bayesian fixed-interval smoothing algorithms in singular state-space systems, MLSP, Sept. 2-4, Grenoble, France, 2009.

[3] C. Andrieu, C. M. Davy, and A. Doucet, Efficient particle filtering for jump Markov systems. Application to time-varying autoregressions, IEEE Trans. Sig. Proc., vol. 51, no. 7, pp. 1762-1770, 2003.

[4] V. S. Barbu and N. Limnios, Semi-Markov chains and hidden semi-Markov models toward applications, Lectures Notes in Statistics, Springer, 2008.

[5] D. Benboudjema and W. Pieczynski, Unsupervised statistical segmentation of non stationary images using triplet Markov fields, IEEE Trans. PAMI, vol. 29, no. 8, pp. 1367-1378, 2007.

[6] R. Bhar and S. Hamori, Hidden Markov Models. Applications to financial economics, Kluwer Academic, 2004.

[7] C. Biernacki, G. Celeux, G. and G. Govaert, Assessing a mixture model for clustering with the integrated complete likelihood. IEEE trans. PAMI, vol. 22, pp.719-725, 2000.

[8] J. A. Bilmes, What HMMs can do, IEICE Trans. on Information and Systems, vol. E89D, no. 3, pp. 869-891, 2006.

[9] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[10] J. Blanchet and F. Forbes, Triplet Markov fields for the classification of complex structure data, IEEE Trans. PAMI, vol. 30, no. 6, pp. 1055 - 1067, 2008.

[11] J. Blanchet and M. Vignes, A model-based approach to gene clustering with missing observation reconstruction in a Markov random field framework, J. Comp. Biol., vol. 16, no.3, pp.475–486, 2009.

[12] C. Bouveyron, S. Girard, and C. Schmid, High dimensional data clustering, Comput. Statist. Data Anal., vol.52, no.1, pp.502-519, 2007.

[13] Y. Boykov, O. Veksler, and R. Zabih, Fast approximate energy minimization via graph cuts, IEEE trans. PAMI, vol.23, no.11, pp.1222–1239, 2001.

[14] N. Brunel and W. Pieczynski, Unsupervised signal restoration using hidden Markov chains with copulas, Signal Processing, vol. 85, no. 12, pp. 2304-2315, 2005.

[15] W. Byrne and A. Gunawardana, Convergence theorems of Generalized Alternating Minimization Procedures, J. Mach. Lear. Res., 2005, vol. 6, pp.2049-2073.

[16] O. Cappé, E. Moulines, and T. Ryden, Inference in hidden Markov models, Springer, 2005.

[17] C. Carincotte, S. Derrode, G. Sicot, and J.-M. Boucher, Unsupervised image segmentation based on a new fuzzy HMC model, ICASSP 2004, May 17-21, Montreal, Canada, 2004.

[18] C. Carincotte, S. Derrode et S. Bourennane, Unsupervised change detection on SAR images using fuzzy hidden Markov chains, IEEE TGRS, vol. 44, no. 2, pp. 432-441, Feb. 2006.

[19] G. Celeux, F. Forbes, C.P. Robert and M. Titterington, Deviance Information Criteria for missing data models, Bayesian Analysis, 2007.

[20] G. Celeux and G. Govaert, Gaussian Parsimonious Clustering Models, Pat. Rec., vol.28, pp.781–793, 1995.

[21] G. Celeux, F. Forbes, and N. Peyrard, EM procedures using mean field-like approximations for Markov model-based image segmentation, Pat. Rec., vol. 36, no.1, pp.131–144, 2003.

[22] O. L. V Costa, M. D. Fragoso, and R. P. Marques, Discrete time Markov jump linear systems, NY, Springer-Verlag, 2005.

[23] T. Crivelli, G. Piriou, B. Frias, et al., Temporal modeling of motion textures with mixed-states Markov chains , ICASSP 09, Las Vegas, USA, March 30 - April 04, 2008.

[24] T. Crivelli, G. Piriou, P. Bouthemy, et al., Simultaneous motion detection and background reconstruction with a mixed-state conditional Markov random field, ECCV 2008.

[25] M.S. Crouse, R. D. Nowak, and R. G. Baraniuk, Wavelet-based statistical signal processing using hidden Markov models, IEEE Trans. Sig. Proc., vol.46, no.4, pp.886-902, 1998.

[26] J.-P. Delmas, An equivalence of the EM and ICE algorithm for exponential family, IEEE Trans. Sig. Proc. vol. 45, no. 10, pp. 2613-2615, 1997.

[27] A. Dempster, N. Laird, and D. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. Roy. Statist. Soc. Ser. B, vol. 39, pp.1–38, 1977.

[28] S. Derrode and W. Pieczynski, Signal and image segmentation using pairwise Markov chains, IEEE Trans. Sig. Proc. vol. 52, no. 9, pp. 2477-2489, 2004.

[29] D. A. Dewasurendra, P. H. Bauer, and K. Premaratne, Evidence Filtering, IEEE Trans. Sig. Proc. vol. 55, no 12, pp. 5796-5805, 2007.

[30] P. Doukhan, G. Oppenheim, and M. S. Taqqu, Long-Range Dependence, Birkhauser, 2003.

[31] J-B. Durand, P. Gonçalves and Y. Guédon, Computational methods for Hidden Markov Tree Models- An application to Wavelet Trees, IEEE, Trans. Sig. Proc., vol. 52, no.9, pp.2551–2560., 2004.

[32] F. Forbes and G. Fort. Combining Monte Carlo and Mean field like methods for inference in hidden Markov Random Fields. IEEE Trans. Im. Proc., vol. 16, no.3, pp.824-837, 2007.

[33] F. Forbes and N. Peyrard, Hidden Markov random field model selection criteria based on mean field-like approximations, IEEE Trans. PAMI, vol. 25, no.8, pp.1089-1101,(46), No. 4, April 1998, pp. 886-902. 2003.

[34] G. Fort and E. Moulines. Convergence of the Monte-Carlo EM for curved exponential families. Ann. Statist. vol.31, n. 4, pp. 1220-1259. 2003.

[35] S. Foucher, M. Germain, J.-M. Boucher, and G. B. Benié, Multisource classification using ICM and Dempster-Shafer theory, IEEE Trans. Instr. Meas., vol. 51, no. 2, pp. 277-281, 2002.

[36] N. de Freitas, P. Hojen-Sorensen, M. Jordan and S. Russel, Variational MCMC. Tech. Rep. UC Berkeley, 2001.

[37] Z. Ghahramani, Hidden Markov models: applications in computer vision, in Hidden Markov models: applications in computer vision, World Scientific, 2001.

[38] P.J. Green, S. Richardson, Hidden Markov models and disease mapping, J. Am. Stat. Assoc. vol.97, pp.1055-1070, 2002.

[39] M. Hatt, F. Lamare, N. Boussion, et al., Fuzzy hidden Markov chains segmentation for volume determination and quantitation in PET, Physics in Medicine and Biology, vol. 52, no. 12, pp. 3467-3491, 2007.

[40] F. V. Jensen, Introduction to Bayesian network, Springer-Verlag New York, 1996.

[41] M.I. Jordan, ed. An introduction to variational methods for graphical models, MIT Press, 1999.

[42] R. Kass and A. Raftery, Bayes factor, J. Amer. Stat. Assoc. vol.90, pp.733-795, 1995.

[43] C. Kervrann, J. Boulanger, T. Pécot, P. Pérez, Discriminant random field and patch-based redundancy analysis for image change detection, MLSP, Sept. 2-4, Grenoble, France, 2009.

[44] T. Koski, Hidden Markov models for bioinformatics, Kluwer Academic Publishers, Netherlands, 2001.

[45] S. Kumar and M. Hebert. Discriminative random fields. Int. J. Comput. Vision, vol.68, no.2, pp.179-201, 2006.

[46] P. Lanchantin, J. Lapuyade-Lahorgue, and W. Pieczynski, Unsupervised segmentation of triplet Markov chains with long-memory noise, Sig. Proc., vol. 88, no. 5, pp. 1134-1151, 2008.

[47] S. Le Cam, F. Salzenstein, and C. Collet, Fuzzy pairwise Markov chain to segment correlated noisy data, Sig. Proc., vol. 88, no. 10, pp. 2526-2541, 2008.

[48] S. Le Cam, C. Collet, and F. Salzenstein, Detection of transient signals: local approach using a Markovian tree with frequency selectivity, MLSP, Sept. 2-4, Grenoble, France, 2009.

[49] L. Liao and T. S. Lin, MR brain image segmentation based on kernelized fuzzy clustering using fuzzy Gibbs Random Field model, IEEE/ICME, Beijing, China, May 23-27, 2007.

[50] R.J. Little and D.B. Rubin, Statistical analysis with missing data, New-York: Wiley, second edition, 2002.

[51] T. Minka, Divergence measures and message passing, Microsoft Research Tech. Rep., MSR-TR- 2005-173, Dec. 7 2005.

[52] A. Mohammad-Djafari and H. Ayasso, Variational Bayes and Mean Field Approximations for Markov field unsupervised estimation, MLSP, Sept. 2-4, Grenoble, France, 2009.

[53] E. Monfrini, Identifiabilité et Méthode des Moments dans les mélanges généralisés de distributions du système de Pearson, PhD (in French), Université Paris VI, 2002.

[54] R. Muoz-Salinas, R. Medina-Carnicer, F.J. Madrid-Cuevas, and A. Carmona-Poyato, Multi-camera people tracking using evidential filters, International Journal of Approximate Reasoning, vol. 50, no. 5, pp. 732-749, May 2009.

[55] R. B. Nelsen, An introduction to copulas, Lecture Notes in Statistics, Vol. 139, Springer, New York, 1998.

[56] W. Pieczynski, Sur la convergence de l'estimation conditionnelle itérative, Comptes Rendu de l'Académie des Sciences-Mathématique, vol. 346, no. 7-8, pp. 457-460, Avril 2008.

[57] W. Pieczynski, Multisensor triplet Markov chains and theory of evidence, Int. J. Approx. Reas., vol. 45, no. 1, pp. 1-16, 2007.

[58] W. Pieczynski and N. Abbassi, Exact filtering and smoothing in short or long memory stochastic switching systems, MLSP, Sept. 2-4, Grenoble, France, 2009.

[59] W. Pieczynski and D. Benboudjema, Multisensor triplet Markov fields and theory of evidence, Image and Vision Computing, vol. 24, no. 1, pp. 61-69, 2006.

[60] W. Pieczynski and F. Desbouvries, On triplet Markov chains, ASMDA 2005, Brest, France, May 2005.

[61] W. Pieczynski and A.-N. Tebbache, Pairwise Markov random fields and segmentation of textured images, Machine Graphics & Vision, vol. 9, no. 3, pp. 705-718, 2000.

[62] E. Ramasso, C. Panagiotakis, D. Pellerin, et al., Human action recognition in videos based on the Transferable Belief Model, Pat. Ana. and Appl., vol. 11, no. 1, pp. 1-19, 2008.

[63] E. Ramasso, Contribution of Belief Functions to Hidden Markov Models with an application to fault diagnostic, MLSP, Sept. 2-4, Grenoble, France, 2009.

[64] B. Ristic, S. Arulampalam, and N. Gordon, Beyond the Kalman Filter - Particle filters for tracking applications, Artech House, Boston, MA, 2004.

[65] C.P. Robert and G. Casella, Monte Carlo Statistical Methods. Springer-Verlag, New York, 1999.

[66] S. Ruan, B. Moretti, J. Fadili, and D. Bloyet, Fuzzy Markovian segmentation in application of magnetic resonance images, Comp. Vis. and Ima. Underst., 85, pp. 54-69, 2002.

[67] D. Rubin, Inference and missing data, Biometrika, vol.63, pp.581–592, 1976.

[68] F. Salzenstein and W. Pieczynski, Parameter estimation in hidden fuzzy Markov random fields and image segmentation, Graph. Mod. and Ima. Proc., vol. 59, no. 4, pp. 205-220, 1997.

[69] F. Salzenstein, C. Collet, S. Le Cam, and M. Hatt, Non stationary fuzzy Markov chains, Pat. Rec. Let., vol. 28, no. 16, pp. 2201-2208, 2007.

[70] B. Scherrer, F. Forbes, C. Garbay and M. Dojat, Distributed local MRF models for tissue and structure brain segmentation, IEEE Trans. Med. Im. To appear, 2009.

[71] G. Schwarz, Estimating the dimension of a model, Ann. Stat., vol.6, pp.461–464, 1978.

[72] G. Sfikas, C. Nikou, C. Heinrich, and N. Galatsanos, On the optimization of probability vector MRFs in image segmentation, MLSP, Sept. 2-4, Grenoble, France, 2009.

[73] G. Shafer, A mathematical theory of evidence, Princeton University Press, Princeton, 1976.

[74] C. Tison, J.M. Nicolas, F. Tupin, and H. Maître, A new statistical model for Markovian classification of urban areas in high-resolution SAR images, IEEE TGRS, vol.42, no.10, pp.2046-2057, 2004.

[75] M. Vignes and F. Forbes, Gene clustering via integrated Markov models combining individual and pairwise features, IEEE Trans. Comp. Bio. Bioinf., vol.6, no.2, pp.260-270, 2009.

[76] T. Vincent, L. Risser, P. Ciuciu, and J. Idier, Robust extrapolation scheme for fast estimation of 3D Ising field partition functions. Application to within-subject fMRI data analysis, MLSP 2009, Sept. 2-4, Grenoble, France, 2009.

[77] C. Wolf, Families of Markov models for document image segmentation, MLSP, Sept. 2-4, Grenoble, France, 2009.

[78] K. Yamazaki, On the likelihood function of HMMs for a long data sequence, MLSP, Sept. 2-4, Grenoble, France, 2009.

[79] J. Zhang, The Mean Fields Theory in EM Procedures for Markov Random Fields. IEEE Trans. sig. proc., vol. 40, n. 10, pp.2570-2583, 1992.