Information Theory & Applications Workshop (ITA), UCSD, La Jolla, Feb. 2009

Localizing, Forgetting, and Likelihood Filtering in State-Space Models

Hans-Andrea Loeliger, Lukas Bolliger, Christoph Reller
ETH Zurich
Dept. of Information Technology and Electrical Engineering,
8092 Zürich, Switzerland
Email: {loeliger, bolliger, reller}@isi.ee.ethz.ch

Sascha Korl
Hilti Corporation
BU Measuring Systems
9494 Schaan, Liechtenstein
Email: sascha.korl@hilti.com

***Abstract*—The context of this paper are cycle-free factor graphs such as hidden Markov models or linear state space models. The paper offers some observations and suggestions on "localizating" such models and their likelihoods. First, it is suggested that a localized version of the model likelihood, which is easily computed by forward sum-product message passing, may be useful for feature extraction and detection. Second, the notion of a "local" model (local factor graph) is introduced. A first class of local models arises from exponential message damping and scale factors as in recursive least squares. A second class of local models arises from the problem of estimating the moment of a model switch from some known model A to some known model B. This problem can be solved by forward sum-product message passing in model A and backward sum-product message passing in model B. It is pointed out that this method is applicable to pulse position estimation for any pulse with a (deterministic or stochastic) state space model.**

I. Introduction

We consider sum-product message passing in cycle-free factor graphs of state space models such as hidden Markov models or linear Gaussian models [1], [2]. In this classical setting, we make some observations and point out some new applications, all of which revolve around some form of "localization".

Specifically, we consider models

$$p(x_0, x_1, \ldots, x_n, y_1, y_2, \ldots, y_n) = p(x_0) \prod_{k=1}^{n} p(x_k, y_k | x_{k-1}) \tag{1}$$

where $X_0, X_1, \ldots, X_n$ are "hidden" variables (state variables) and $Y_1, \ldots, Y_n$ are observable variables. The factor graph of (1) is shown in Fig. 1. We use Forney-style factor graphs where nodes / boxes represent factors and edges represent variables [1], [2].

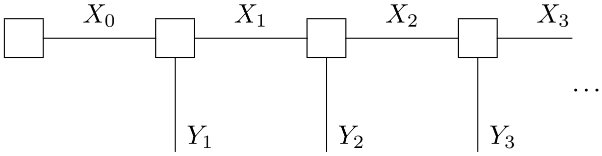

Fig. 1. Factor graph (Forney-style) of state space model.

We are mainly interested in the case where n is large. We will use the notation of discrete variables, but the translation to continuous variables is straightforward.

In section II, we review several ways to compute the model likelihood $p(y_1, \ldots, y_n)$, a "localization" of which is proposed in section III. In section IV, we consider exponential message damping and scale factors as in RLS (recursive least squares) algorithms, which leads us to the notion of a "local" model (local factor graph). Another class of local models appears in Section V, where we show that the position of a model change—the position of a switch from some known model A to some known model B—can be estimated from forward sum-product message passing in (the factor graph of) model A and backward sum-product message passing in (the factor graph of) model B. Finally, we point out that this method can be used for pulse position estimation for any pulse with a determistic or stochastic state space model.

II. Computing the Model Likelihood

Let us recall the computation of $p(y_1, \ldots, y_n)$ for given observations $y_1, \ldots, y_n$. In a first version, we have

$$p(y_1, \ldots, y_n) = \sum_{x_n} \sum_{x_0, \ldots, x_{n-1}} p(x_0, \ldots, x_n, y_1, \ldots, y_n) \tag{2}$$

$$= \sum_{x_n} \overrightarrow{\mu}_{X_n}(x_n), \tag{3}$$

where $\overrightarrow{\mu}_{X_n}$ is the forward (left-to-right) sum-product message along the edge X_n. This method has probably been used most often.

In a second version, we have

$$p(y_1, \ldots, y_n) = \sum_{x_{k-1}} \sum_{x_k} \overrightarrow{\mu}_{X_{k-1}}(x_{k-1}) p(x_k, y_k | x_{k-1}) \overleftarrow{\mu}_{X_k}(x_k) \tag{4}$$

$$= \sum_{x_k} \overrightarrow{\mu}_{X_k}(x_k) \overleftarrow{\mu}_{X_k}(x_k) \tag{5}$$

where $\overleftarrow{\mu}_{X_k}$ is the backward sum-product message along X_k. This method will be used in section V.

In actual implementations, proper scaling of the messages $\overrightarrow{\mu}_{X_k}$ and $\overleftarrow{\mu}_{X_k}$ is required for numerical stability. The computation of $p(y_1, \ldots, y_n)$ by means of (3)–(5) then amounts to

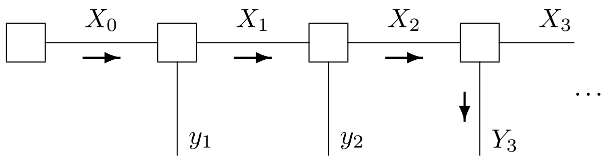

Fig. 2. Computing $p(y_k|y_1,\ldots,y_{k-1})$ (for $k=3$) for known $Y_1=y_1$, $\ldots$, $Y_{k-1}=y_{k-1}$ (but unknown $Y_k, Y_{k+1},\ldots$) according to (8).

keeping track of the scale factors (preferably in the logarithmic domain), cf. [3].

In a third version, we have

$$\log p(y_1,\ldots,y_n) = \log p(y_1) + \sum_{k=2}^{n} \log p(y_k|y_1,\ldots,y_{k-1}) \tag{6}$$

and

$$p(y_k|y_1,\ldots,y_{k-1}) \propto \sum_{x_0,\ldots,x_k} p(x_0,\ldots,x_k,y_1,\ldots,y_k) \tag{7}$$

$$= \overrightarrow{\mu}_{Y_k}(y_k) \tag{8}$$

as is illustrated in Fig. 2. Note that the missing scale factor in (8) can be recovered from $\sum_{y_k} p(y_k|y_1,\ldots,y_{k-1}) = 1$. This method will be used in the next section.

III. Signal Class Likelihood Filter

A "localized" version of the model log-likelihood (6) is obtained by defining the instantaneous log-likelihood

$$L_k = \gamma L_{k-1} + \log p(y_k|y_1,\ldots,y_{k-1}) \tag{9}$$

for some positive real $\gamma < 1$ (typically $\gamma \approx 1$). The computation of $L_1, L_2, \ldots$ from $y_1, y_2, \ldots$ using (8) and (9) may be viewed as a nonlinear filter.

For example, consider questions of the type "Is the telephone ringing"? How likely are the recent observations $y_n, y_{n-1}, \ldots$ up to time n (= now) under some given signal model? Note that the (backwards) time horizon has not been specified. (The telephone may have been ringing for a minute or just for a second.) A practical solution to such detection problems may be obtained by testing the instantaneous log-likelihood L_n against some threshold.

(Note that there is no claim for optimality—there is not even a fully defined problem. Optimal solutions to similar, but well-defined problems are available in the literature on "quickest detection" [4].)

More generally, such "signal class likelihood filters" may be used for model-based feature extraction. A two-sided version of (9) may be used for off-line applications.

IV. Forgetting Factors and Local Models

The "damping" of messages by raising them to some power $\gamma < 1$ ($\gamma > 0$) is a practical device that has been used in many applications. In particular, if some message $\overrightarrow{\mu}_X(x)$ is a Gaussian probability distribution with covariance matrix V, then $\overrightarrow{\mu}_X(x)^\gamma$ is a Gaussian distribution with covariance matrix V/γ. In this way, the scale factor in RLS (recursive least squares) algorithms may be viewed as an example of such exponential message damping [2].

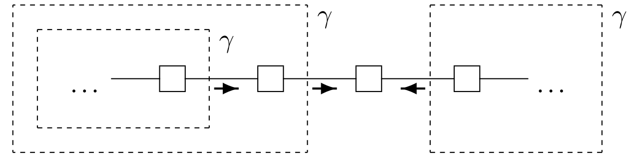

Fig. 3. Recursive damping of remote factors.

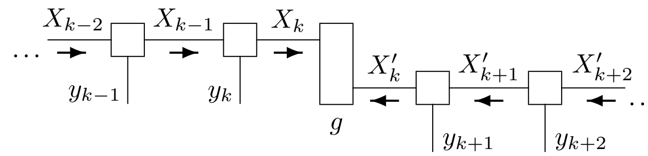

Fig. 4. Factor graph with a model switch.

Sum-product message passing with such damping may be viewed as exact (undamped) sum-product message passing in a modified model / factor graph as in Fig. 3, which expresses decreasing confidence in, or reliance on, remote parts of the model.

An unusual feature of this modified factor graph is that it depends on the variable / edge in the focus; each variable, when focussed on, has its own "local" model. Another example of such a local model is given in the next section.

V. Detection of Model Switches and Pulse Positions

Consider a situation in which $y_1,\ldots,y_k$ are generated by some given model A while $y_{k+1},\ldots,y_n$ are generated by some given model B. Assume that we do not know the moment k of the model switch and we wish to estimate it.

Let H_k be the hypothesis that the model switch occurs between the observations y_k and y_{k+1}. Then

$$\begin{aligned} p(y_1,\ldots,y_n|H_k) &= \sum_{x_0,\ldots,x_k} \sum_{x'_k,\ldots,x'_n} p(x_0,\ldots,x_k,y_1,\ldots,y_k|A) \\ &\quad g(x_k,x'_k)\, p(x'_k,\ldots,x'_n,y_{k+1},\ldots,y_n|B) \end{aligned} \tag{10}$$

where the "glue factor" $g(x_k, x'_k)$ expresses the constraints on the final state X_k of model A and the inital state X'_k of model B. The factor graph of (10) is shown in Fig. 4. Based on this factor graph, the likelihood $p(y_1,\ldots,y_n|H_k)$ can be computed according to (4) or (5) by forward message passing in model A and backward message passing in model B as indicated in Fig. 4.

Note that $p(y_1,\ldots,y_n|H_k)$ can be computed for all k simultaneously by a single forward sum-product sweep in model A and a single backward sum-product sweep in model B; only the final computation according to (4) or (5) needs to be done individually for each k.

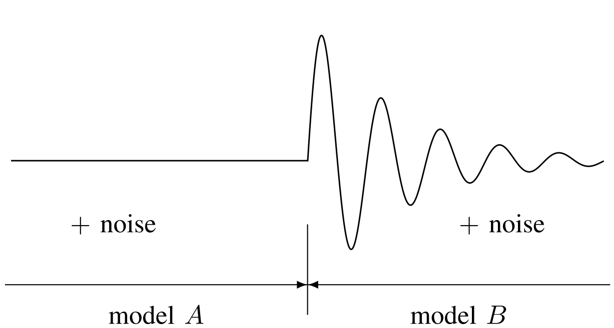

Fig. 5. Pulse position estimation: one-sided pulse.

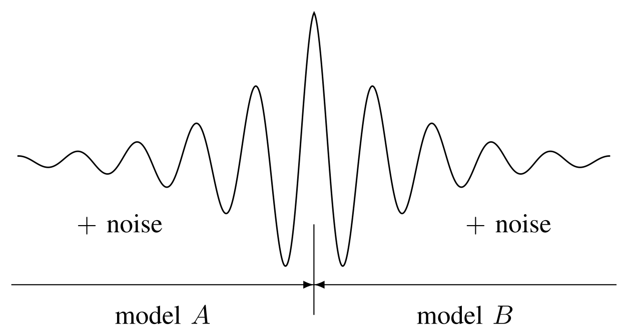

Fig. 6. Pulse position estimation: two-sided pulse.

As an example of such a problem, consider the estimation of the unknown position of a pulse as in Fig. 5 or 6 in additive white Gaussian noise (AWGN). In Fig. 5, model A is white Gaussian noise and model B is a second-order linear system observed with AWGN; in Fig. 6, both model A and model B are second-order linear systems observed with AWGN. In both cases, the glue factor $g(x_k, x_k')$ is required to enforce the proper initial / final conditions of the linear systems. (Sum-product message passing through these linear systems amounts to Kalman filtering [2], but the message computation tables in [2] neglect the scale factors that are here required; details will be given elsewhere.) More generally, we note that this method for pulse position estimation can be used for any pulse-with-noise model in state space form.

References

[1] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Proc. Mag.,* Jan. 2004, pp. 28–41.

[2] H.-A. Loeliger, J. Dauwels, Junli Hu, S. Korl, Li Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proceedings of the IEEE,* vol. 95, no. 6, pp. 1295–1322, June 2007.

[3] D. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3498–3508, August 2006.

[4] H. V. Poor and O. Hadjiliadis, *Quickest Detection.* Cambridge University Press, 2009.