MCMC INFERENCE OF THE SHAPE AND VARIABILITY OF TIME-RESPONSE SIGNALS

Dmitriy A. Katz-Rogozhnikov, Kush R. Varshney, Aleksandra Mojsilović, and Moninder Singh

Business Analytics and Mathematical Sciences Department, IBM Thomas J. Watson Research Center
1101 Kitchawan Rd., Route 134, Yorktown Heights, NY 10598, USA

ABSTRACT

Signals in response to time-localized events of a common phenomenon tend to exhibit a common shape, but with variable time scale, amplitude, and delay across trials in many domains. We develop a new formulation to learn the common shape and variables from noisy signal samples with a Bayesian signal model and a Markov chain Monte Carlo inference scheme involving Gibbs sampling and independent Metropolis-Hastings. Our experiments with generated and real-world data show that the algorithm is robust to missing data, outperforms the existing approaches and produces easily interpretable outputs.

***Index Terms*—** time-response signal, multiple alignment, Markov chain Monte Carlo, outsourcing

1. INTRODUCTION

Responses over time to singular events are studied in many domains, for example, the electric potential in the brain of a subject who has been presented with a visual stimulus [1], the financial performance of a company that has agreed to an outsourcing engagement [2], and the vital signs of a patient who has been administered a medication. Such signals tend to exhibit a common 'shape' or structure, but show variability in amplitude, delay, and time scale across subjects, companies, or patients. Given a set of noisy signal samples, we would like to infer the common shape and also understand the variability associated with shifts and scalings of time and amplitude.

The general signal model that we focus on is:

$$r_i(t) = A_i f(b_i t + d_i) + \text{noise}(t), \quad i = 1, \ldots, n, \tag{1}$$

where $r_i(t)$ is the time-response signal to an event for trial i, $f(t)$ is the common shape of the signals, and A_i, b_i and d_i are amplitude scale, time scale and time shift parameters. Given $\{r_1(t), \ldots, r_n(t)\}$, the general goal is to understand the nature of the time response by determining $f(t)$ and characterizing the variability of A, b and d.

The problem considered here is different from that considered in dynamic time warping, see [3] and references therein, in two ways. First, we consider signals with large amounts of additive noise, whereas dynamic time warping is best suited to signals with little to no noise. Second, we are aiming for interpretability through a specific parameterization with A, b and d, which is not the case with dynamic time warping.

The work of Listgarten et al., which considers a hierarchical Bayesian approach for the alignment of several time series, works well with noisy signals including the type considered in this paper [4, 5]. The latent trace that is learned in that work has a similar role as $f(t)$ in (1). However, the other model parameters learned in that work are difficult to interpret in comparison to the three scale and shift parameters in (1); also nonuniform scalings of time are produced, which may be undesirable in many applications.

Additionally, dynamic time warping and the approach of Listgarten et al. are not applicable when there are missing data values in the measured time-response signals. In light of shortcomings of existing approaches such as limited robustness to noise and missing values, nonuniform time scalings in solutions, and difficult interpretability of solution parameters, we propose a new inference approach built around a hierarchical Bayesian extension of the signal model (1). To make the model tractable, we take $f(t)$ to be a spline function. Such a Bayesian signal processing model, with random variables for shifts and scalings of time and amplitude as well as a common spline function, has not been considered in the literature before. We formulate a Markov chain Monte Carlo (MCMC) method incorporating Gibbs sampling and the Metropolis-Hastings algorithm for inference [6].

Simply averaging the n signals at each time has been a standard technique in neurophysiology for decades [1], and is used elsewhere as well. We compare the proposed method to simple averaging and the method of Listgarten et al. [4], and find the proposed method to be superior to these other two approaches on data generated from the model. Moreover, we find the proposed method to perform better on real-world data from a business analytics application (see Sec. 5). We focus on the business analytics application due to space constraints; others, such as the aforementioned neurophysiology application may also be considered [1].

2. HIERARCHICAL BAYESIAN SIGNAL MODEL

As discussed in Sec. 1, the general signal model that we consider is given by (1). The particular noise we consider is zero-mean white Gaussian noise with variance σ^2. The whiteness of the noise in the model introduces statistical independencies among times which we take advantage of in inference. The particular form of the common shape $f(t)$ we consider is a piecewise polynomial spline with a fixed number of knots m. Let the times of the knots be $t_1, t_2, \ldots, t_m$, and the values of the common shape function at those times be $f_1, f_2, \ldots, f_m$. The value $f(\tilde{t})$ at a time point $\tilde{t}$ other than the knots may be interpolated given $f_1, \ldots, f_m$. The fixed parameterization for $f(t)$ in terms of $f_1, \ldots, f_m$ permits us to develop a tractable inference procedure. The signal $r_i(t)$ is measured at l_i (not necessarily uniformly spaced) time points $t_1^{(i)}, \ldots, t_{l_i}^{(i)}$ and is denoted $r_1^{(i)}, \ldots, r_{l_i}^{(i)}$. The number of measured points and their times may be different for different trials, and different from the common shape function.

There are $3n + m$ random variables in the model we would like to infer, namely (A_i, b_i, d_i), $i = 1, \ldots, n$ and f_j, $j = 1, \ldots, m$. Uninformative priors are taken for the f_j. The priors of the scale and shift variables, $p(A_i; \theta_A)$, $p(b_i; \theta_b)$ and $p(d_i; \theta_d)$ with hyperpa-

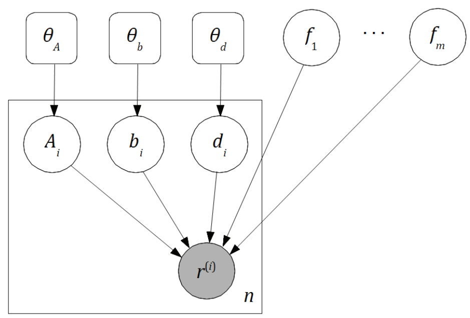

Fig. 1. Graphical model representation of proposed hierarchical Bayesian signal model.

rameters θ_A, θ_b and θ_d, are the same for all i and are given later in the paper. A graphical model representation of the signal model is provided in Fig. 1. Plate notation is used in the figure to indicate repetition of the variables for each trial, $i = 1, \ldots, n$. Let $\mathbf{A}$ denote all the variables $A_1, \ldots, A_n$, $\mathbf{b}$ denote all the variables $b_1, \ldots, b_n$, and $\mathbf{d}$ denote all the variables $d_1, \ldots, d_n$.

Due to the additive Gaussian noise, the conditional pdf of f_j,

$$p\left(f_j|\mathbf{A}, \mathbf{b}, \mathbf{d}, f_1, \ldots, f_{j-1}, f_{j+1}, \ldots, f_m, r_1^{(1)}, \ldots, r_{l_n}^{(n)}\right),$$

is Gaussian. The mean and variance of this Gaussian may be derived based on spline interpolation formulas. For example with a piecewise linear $f(t)$, letting

$$\begin{aligned}
\mathcal{T}_{ij}^- &= \{k \mid t_{j-1} < b_i t_k^{(i)} + d_i < t_j\}, \\
\mathcal{T}_{ij}^+ &= \{k \mid t_j < b_i t_k^{(i)} + d_i < t_{j+1}\}, \\
\tau_{ijk}^- &= \tfrac{b_i t_k^{(i)} + d_i - t_j}{t_{j-1} - t_j}, \\
\tau_{ijk}^+ &= \tfrac{b_i t_k^{(i)} + d_i - t_j}{t_{j+1} - t_j},
\end{aligned}$$

the mean of the Gaussian is

$$\frac{\sum_{i=1}^n \left(\sum_{\mathcal{T}_{ij}^-} r_k^{(i)} - A_i f_{j-1} \tau_{ijk}^- + \sum_{\mathcal{T}_{ij}^+} r_k^{(i)} - A_i f_{j+1} \tau_{ijk}^+\right)}{\sum_{i=1}^n \left(\sum_{\mathcal{T}_{ij}^-} |A_i|(1 - \tau_{ijk}^-) + \sum_{\mathcal{T}_{ij}^+} |A_i|(1 - \tau_{ijk}^+)\right)}, \tag{2}$$

and the variance is

$$\frac{\sigma^2}{\sum_{i=1}^n \left(\sum_{\mathcal{T}_{ij}^-} (A_i(1 - \tau_{ijk}^-))^2 + \sum_{\mathcal{T}_{ij}^+} (A_i(1 - \tau_{ijk}^+))^2\right)}. \tag{3}$$

With a linear spline, f_j depends on f_{j-1} and f_{j+1}, but not the other values of the shape function; this may or may not be true with other interpolating splines.

A_i depends on the measurements, b_i and d_i of trial i as well as the common f_j values, but not on the measurements, b_i and d_i of any other trial $i' \neq i$. Similarly for b_i and d_i. The conditional densities of the scale and shift variables, due to the white Gaussian noise, are:

$$p\left(A_i \mid b_i, d_i, f_1, \ldots, f_m, r_1^{(i)}, \ldots, r_{l_i}^{(i)}\right) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{l_i} \left(r_k^{(i)} - A_i f\left(b_i t_k^{(i)} + d_i\right)\right)^2\right), \tag{4}$$

$$p\left(b_i \mid A_i, d_i, f_1, \ldots, f_m, r_1^{(i)}, \ldots, r_{l_i}^{(i)}\right) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{l_i} \left(r_k^{(i)} - A_i f\left(b_i t_k^{(i)} + d_i\right)\right)^2\right), \tag{5}$$

and

$$p\left(d_i \mid A_i, b_i, f_1, \ldots, f_m, r_1^{(i)}, \ldots, r_{l_i}^{(i)}\right) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{l_i} \left(r_k^{(i)} - A_i f\left(b_i t_k^{(i)} + d_i\right)\right)^2\right). \tag{6}$$

In (4)–(6), the $f(b_i t_k^{(i)} + d_i)$ terms involve spline interpolation.

3. MCMC SAMPLING

The new hierarchical Bayesian signal model set forth in the previous section factors in such a way (illustrated in Fig. 1) that lends itself to inference via Gibbs sampling [6]. Each of the $3n + m$ random variables is sampled in turn, conditioned on all of the other variables and the measured time-response signals.

The f_j are sampled within a Gibbs iteration according to

$$p\left(f_j|\mathbf{A}, \mathbf{b}, \mathbf{d}, f_1, \ldots, f_{j-1}, f_{j+1}, \ldots, f_m, r_1^{(1)}, \ldots, r_{l_n}^{(n)}\right),$$

which are Gaussians with means and variances discussed in the previous section.

The A_i, b_i and d_i are difficult to sample directly due to the spline interpolation. For these variables, we use independent Metropolis-Hastings sampling [7, 8]. That is, the variable is drawn independently from its prior distribution and then accepted or rejected according to the conditional densities (4)–(6). For example, at the current Gibbs iteration, a candidate amplitude $A_i^{(candidate)}$ is drawn from $p(A_i; \theta_A)$. $A_i^{(candidate)}$ is accepted as $A_i^{(current)}$ with probability $\min\{R, 1\}$, and $A_i^{(previous)}$ is taken as $A_i^{(current)}$ with probability $\max\{1 - R, 0\}$, where R is the Hastings ratio

$$\frac{\exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{l_i} \left(r_k^{(i)} - A_i^{(candidate)} f\left(b_i t_k^{(i)} + d_i\right)\right)^2\right)}{\exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{l_i} \left(r_k^{(i)} - A_i^{(previous)} f\left(b_i t_k^{(i)} + d_i\right)\right)^2\right)}.$$

Since only the ratio of two conditional densities is needed, the normalizing constant of the conditional pdf is not required. Independent Metropolis-Hastings is less commonly used than perturbation-based and other flavors of Metropolis-Hastings, but in the overall problem at hand, we observe that the independent flavor works well, with short mixing times and without getting 'stuck.'

4. RESULTS ON GENERATED DATA

In this section, we illustrate the proposed signal processing technique on data generated from the model (1). Specifically, we use A_i that are uniform over the interval $[0.5, 1.5]$, b_i that are uniform over the interval $[0.5, 1]$, and d_i that are uniform over the interval $[-3, 3]$. The common shape is $f(t) = (t^3/10 - t^2/2 - t/2)e^{-t/4}$, the zero-mean additive white Gaussian noise has variance $\sigma^2 = 1$, and measurements are taken for all n trials at times $t = 1, 2, \ldots, 16$. Note that in generating the data, $f(t)$ is not a spline; however, we use a piecewise linear $f(t)$ in the MCMC inference. One realization of the data for $n = 100$ is shown in Fig. 2(a).

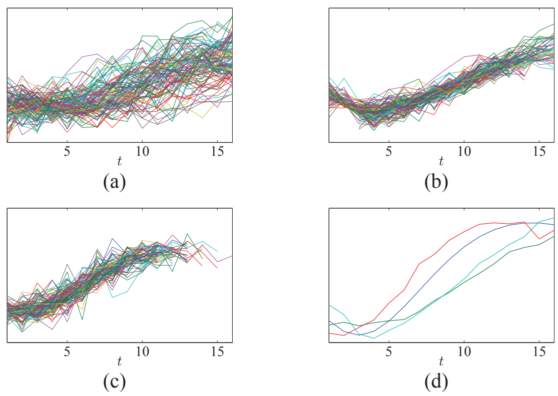

Fig. 2. Generated data (a) response signals, (b) aligned with [4], (c) aligned with proposed MCMC inference, and (d) the averages of (a) in green, (b) in cyan, and (c) in red along with the true $f(t)$ in blue.

	$n = 50$	$n = 100$	$n = 200$	$n = 400$
simple avg	2.11 (0.63)	2.05 (0.42)	2.00 (0.32)	1.98 (0.27)
Listgarten	1.67 (0.60)	1.58 (0.44)	1.51 (0.38)	1.53 (0.33)
HB MCMC	1.24 (1.28)	0.51 (0.41)	0.37 (0.24)	0.24 (0.14)

Table 1. Mean-squared error between true f and estimate for different n averaged over 50 instances of data with standard deviation over the instances in parentheses.

We compare simple averaging, the method of [4], and the proposed MCMC method. The response signals are aligned based on parameters that are learned using both [4] and the proposed method, as shown in Fig. 2(b)–(c), with the averages of those aligned signals shown in Fig. 2(d). The proposed method has the smallest mean-squared error, whereas simple averaging has the largest. Mean-squared error results are presented in Table 1 for the three approaches for different values of n. The proposed MCMC method is superior across all values of n and significantly improves as we increase n. Simple averaging is the worst, and the method of Listgarten et al. is only slightly better.

The proposed MCMC method is also able to handle missing values in the data, by just treating a signal with missing data as one with smaller l_i. The method of Listgarten et al. is not amenable to missing data. In Table 2, we give error results for different percentages of missing data, sampled uniformly over all n and all times. The performance of the proposed MCMC method does degrade as the amount of missing data increases, but this degradation is minimal, showing robustness of the algorithm.

5. RESULTS ON REAL-WORLD DATA

In this section we consider the use of our method in a business analytics application. Corporate leaders are often interested in quantifying an impact of a major initiative to company performance. One such initiative is outsourcing, where an external vendor manages a portion of the company's operations, in order to reduce expenses and increase earnings. Therefore, financial performance metric time signals, such as selling, general and administrative expenses (SG&A) growth rate, and earnings before tax (EBT) growth rate, are expected to respond to a company entering into an outsourcing engagement

	0%	10%	20%	50%
simple avg	2.05 (0.42)	2.06 (0.43)	2.08 (0.42)	2.07 (0.44)
Listgarten	1.58 (0.44)	—	—	—
HB MCMC	0.51 (0.41)	0.53 (0.41)	0.54 (0.45)	0.61 (0.52)

Table 2. Mean-squared error between true f and estimate for different missing data percentages at $n = 100$ averaged over 50 instances of data with standard deviation over the instances in parentheses.

[2]. We use the proposed signal model and inference methodology to understand the impact of outsourcing on the future performance of companies.

We expect outsourcing to change the growth rate of SG&A and EBT for a period of time, with the impact taking a common shape across companies. We also expect the amplitude of the impact to be different and for the impact to occur on different time scales for different companies. In addition, the data available for outsourcing events is the date at which the outsourcing engagement deal was signed, but not when the outsourcing was actually rolled out. It is thus important to include time delay parameters in the signal model. Performance metric measurements are quite noisy; the white Gaussian model is not inappropriate in this setting and is applied because of its mathematical convenience.

We use a uniform distribution over the interval $[0, 1]$ as $p(A_i; \theta_A)$ and a uniform distribution over the interval $[1/2, 1]$ as $p(b_i; \theta_b)$. The delay prior is the following:

$$p(d_i; \theta_d) = \begin{cases} 1/2, & 0 < d_i \leq 1 \\ -d_i/4 + 3/4, & 1 < d_i \leq 3 \\ 0, & \text{otherwise} \end{cases}$$

where time is measured in quarters of years. The $p(d_i; \theta_d)$ distribution is easily sampled using Smirnov transformation of uniform random numbers.

We show results of MCMC inference for SG&A on $n = 249$ companies in Fig. 3 and for EBT on $n = 216$ companies in Fig. 4. Companies whose available response signals contain missing values are discarded to allow comparison with [4], which is why n is different for the two financial metrics. Subfigure (a) shows the n noisy signal samples and (b) shows the result of taking a simple average over the n signals at each time. Subfigure (c) shows the latent trace obtained using [4], whereas (d) shows the median values of 500 MCMC samples of f from iterations 2501 to 3000.[1]

For SG&A, we see that the simple average, the latent trace, and the median common shape $f(t)$ are all similar looking, notably decreasing for about four quarters after $t = 0$. Expenses do decrease as a result of outsourcing. Earnings are also impacted by outsourcing, but less directly and with some delay. The simple average produces a fairly noisy signal for EBT while the latent trace and median common shape are smoother. In all three plots, we see that after $t = 0$, the EBT growth rate initially decreases for a couple of quarters and then increases for a few quarters. The simple average and the median common shape are more in line with each other than the latent trace of [4], which is a result of [4] incorporating nonuniform scalings of time in order to align signals. In the latent trace, the initial decrease in EBT growth rate after $t = 0$ is stretched out in comparison with other parts of the signal. In understanding the business impact of outsourcing, such nonuniformities make interpretation difficult.

[1] Fig. 3(h) contains n curves which are moving window averages of A_i over 500 MCMC samples, illustrating that the Markov chain has sufficiently mixed at iterations 2501 to 3000.

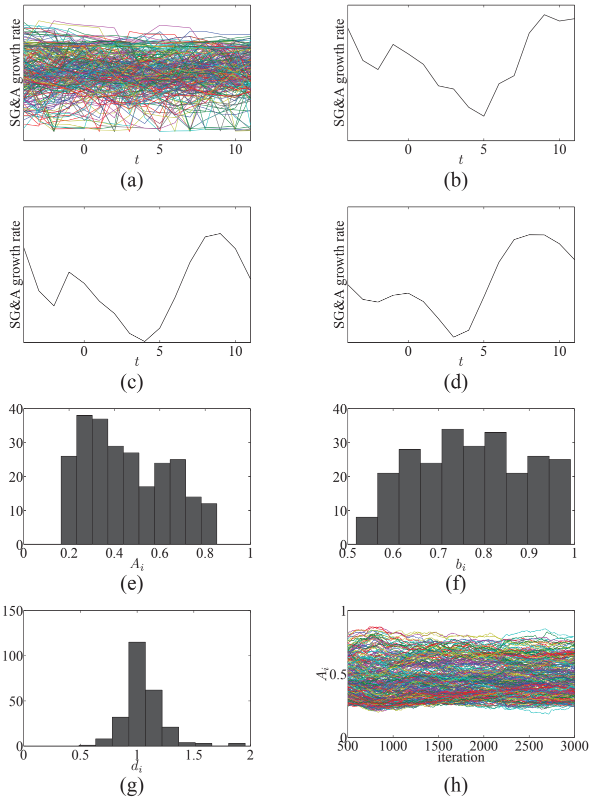

Fig. 3. SG&A (a) time-response signals, (b) simple average, (c) latent trace [4], (d) median common shape $f(t)$, (e) histogram of median amplitudes, (f) histogram of median time scales, (g) histogram of median delays, and (h) moving window average of amplitude MCMC samples.

Another aspect of the hierarchical Bayes model is that we have access to the highly interpretable A_i, b_i, and d_i variables. The figures show histograms over the n companies of median A_i, b_i, and d_i, illustrating the variability of these parameters. Interestingly, the distribution of the delay parameter d_i is approximately the same for both SG&A and EBT, centered around one quarter after the signing of the outsourcing deal; this indicates that the delay parameter truly is capturing the time between the signing of the deal and the roll out. The amplitude distributions are different, which is to be expected because outsourcing has different relative effects on SG&A and EBT for different companies.

6. CONCLUSION

In this paper, we have developed a Bayesian inference methodology for determining the common structure and variability of a collection of time-response signals. It is shown to work well on generated and real-world data, to be robust to missing data, and to produce easily interpretable outputs including uniform time scalings.

That we see a common delay distribution between SG&A and EBT suggests a direction for future work: extending the model to jointly consider more than one signal per trial with common delays, but possibly different amplitudes and time scales. Another extension

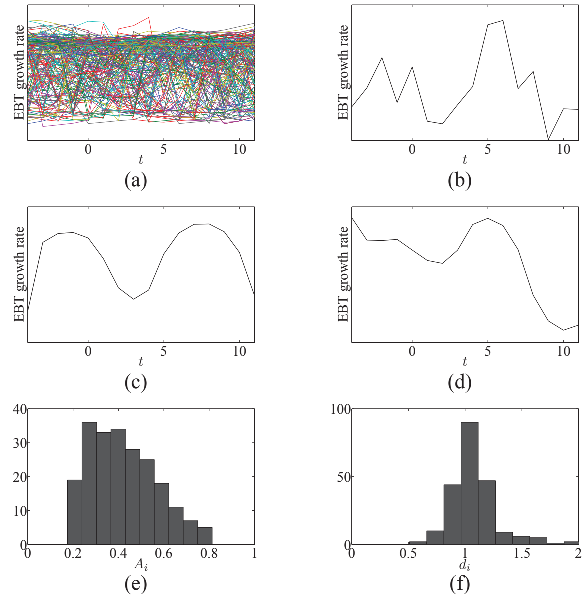

Fig. 4. EBT (a) time-response signals, (b) simple average, (c) latent trace [4], (d) median common shape $f(t)$, (e) histogram of median amplitudes, and (f) histogram of median delays.

of the model includes adding one further layer of hierarchy, by not taking θ_A, θ_b, and θ_d to be fixed hyperparameters, but to be random variables that must also be inferred. In such an extended model, we could understand the variability of A, b, and d in a more rigorous fashion than examining histograms such as those in Fig. 3 and Fig. 4.

7. REFERENCES

[1] R. Quian Quiroga and H. Garcia, "Single-trial event-related potentials with wavelet denoising," *Clin. Neurophysiol.*, vol. 114, no. 2, pp. 376–390, Feb. 2003.

[2] "Business impact of outsourcing: A fact-based analysis," IBM Corp., Armonk, NY, Jan. 2010.

[3] K. Wang and T. Gasser, "Alignment of curves by dynamic time warping," *Ann. Stat.*, vol. 25, no. 3, pp. 1251–1276, Jun. 1997.

[4] J. Listgarten, R. M. Neal, S. T. Roweis, and A. Emili, "Multiple alignment of continuous time series," *Adv. Neural Inf. Process. Syst. 17*, pp. 817–824, 2005.

[5] J. Listgarten, R. M. Neal, S. T. Roweis, R. Puckrin, and S. Cutler, "Bayesian detection of infrequent differences in sets of time series with shared structure," *Adv. Neural Inf. Process. Syst. 19*, pp. 905–912, 2007.

[6] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge, UK: Cambridge University Press, 2003.

[7] L. Tierney, "Markov chains for exploring posterior distributions," *Ann. Stat.*, vol. 22, no. 4, pp. 1701–1728, Dec. 1994.

[8] J. S. Liu, "Metropolized independent sampling with comparisons to rejection sampling and importance sampling," *Stat. Comput.*, vol. 6, no. 2, pp. 113–119, Jun. 1996.