Extracting Motifs from Time Series Generated by Concurrent Activities

Jagannadan Varadarajan[1,2], **Rémi Emonet**[1], **Jean-Marc Odobez**[1,2]
[1]Idiap Research Institute, CH-1920 Martigny, Switzerland.
[2]Ecole Polytechnique Federal de Lausanne, CH-1015 Lausanne, Switzerland.
{vjagann,remi.emonet,odobez}@idiap.ch

Abstract

In this article, we present a model for un-supervised extraction of motifs from multivariate time series. We consider a particular kind of time series where observed values are caused by the superposition of multiple phenomena occuring concurrently and with no synchronization.

1 Introduction and Context

Finding recurrent motifs automatically in time series has been an active research area for many years. Various kind of time series has been considered for motif mining. Discrete time series or regularly sampled continuous time series has been considered for motif detection, both for univariate time series as in [1] or multivariate as in [2]. Spatial motif discovery in DNA sequence has also been deeply studied as in [3].

There are some multivariate time series that are not well managed by the existing methods. These time series are specific in the sense that the observed values are caused by the superposition of multiple phenomena. These phenomena happen concurrently and are not necessarily synchronized.

To illustrate these particular time series, we can take the toy example of a bivariate time series: the overall electric and water consumptions of a house or building. We could expect to spot activity patterns (motifs) like a short high water consumption followed by short high electric consumption (someone filling and then starting a boiler). We could also expect patterns like one hour long alternating water and electric consumption (for a washing machine). Due to the presence of multiple persons, multiple occurrences of these two motifs can occur at the same time and with no particular synchronization. Note that the two motifs share the same basic actions but differ only in the order of execution. Another illustration is the spatio-temporal brain activity data as studied in [4].

We consider this specific kind of time series for which, to our knowledge, existing methods are not well adapted. Our main application domain is activity recognition from video data. With no supervision, we want to extract motion activity patterns (e.g. car passing, pedestrian crossing) to do tasks like abnormality detection and event detection. Here, the considered time series are the amount of motion taken at different positions and orientations in the images. In this context, many research effort such as [5] or [6] concentrated on adapting topic models (originally used for textual document analysis) to the temporal aspect of these data. These approaches still suffer from a inadequate support for non synchronized phenomena.

In this paper, we present a method called Probabilistic Latent Sequential Motif (pLSM). Our pLSM automatically and accurately finds the temporal patterns in multivariate time series featuring superposed phenomena. This is achieved by jointly estimating the motifs (temporal patterns) and when they occur in the observed documents.

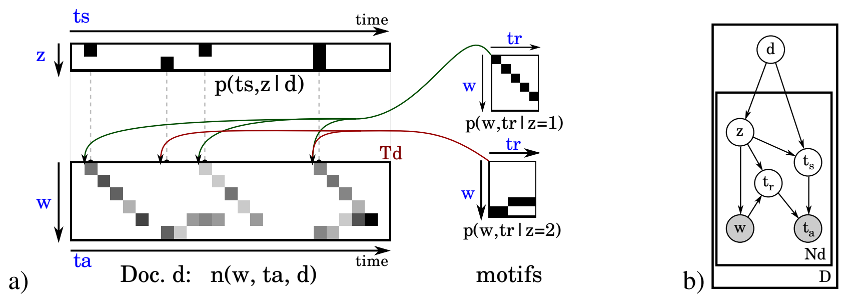

Figure 1: Generative process. a) building a document is done by stochastically pasting each motifs $p(w, tr|z)$ at the positions indicated by the starting-time table $p(ts, z|d)$ of the document (details in section 2). b) graphical model in plate notation.

2 Model: Probabilistic Latent Sequential Motif

Figure 1a illustrates how a document d (among all D documents) is generated from our model. Each document spans T_d discrete time steps and contains words from a vocabulary $V = \{w_i\}_{i=1}^{N_w}$. A document is described by its count matrix $\mathrm{n}(w, t_a, d)$ indicating the number of times a word w occurs at the absolute time t_a within the document.

All documents are generated from a same set of motifs $\{z_i\}_{i=1}^{N_z}$, each defined by $p(w, t_r|z)$. t_r denotes the relative time at which a word occurs within a motif. Within a document, the motifs can start at any time instant. In our model, we capture these start times in $p(t_s, z|d)$. Qualitatively, documents are generated by sampling words in the motifs and placing them in the document relatively to a sampled starting time, as shown in Fig. 1a (see [7] for details). By explicitly modeling $p(w, t_r|z)$ and $p(t_s, z|d)$, we jointly infer the motifs and when they occur. If either the motifs or their occurrence time were fixed, it would be easier to do inference of the other. However, this joint estimation enables to properly handle the superposition of unaligned motif occurrences in documents.

3 Model Inference and Experiments

The proposed model allows an Expectation Maximization algorithm to be used for inference. Inference can be used for two purposes: motif mining, i.e. from a set of learning documents, we infer both the temporal motifs and when they occur; and, motif detection, i.e. for a new document, we use the learnt motifs (we keep them fixed) and only infer when they occur within the document.

Experiments on both synthetic temporal documents and documents extracted from video recordings have shown the effectiveness of the proposed model. Considering the limited space and for better illustration, we redirect the interested reader to [7] and to `http://www.idiap.ch/~remonet/sup/nips10ts/ts.html` for visual illustrations.

Acknowledgments: The authors gratefully acknowledge the financial support from the Swiss National Science Foundation (Project: FNS-198,HAI) and from the 7th framework program of the European Union (Integrated project VANAHEIM(248907) and Network of Excellence PASCAL2)

References

[1] Y. Lin, M. D McCool, and A. A Ghorbani. Motif and anomaly discovery of time series based on subseries join. *Proc. of the Int. MultiConference of Engineers and Computer Scientists*, 2010.

[2] Yuan Li and Jessica Lin. Approximate variable-length time series motif discovery using grammar inference. In *Proc. of the ACM Tenth Int. Workshop on Multimedia Data Mining*, 2010.

[3] J. Buhler and M. Tompa. Finding motifs using random projections. *Journal of computational biology*, 9(2):225242, 2002.

[4] P. Ciuciu, T. Vincent, L. Risser, and S. Donnet. A joint detection-estimation framework for analysing within-subject fMRI data. *Journal de la Société Française de Statistique*, 2010.

[5] Hanna M. Wallach. Topic modeling: beyond bag-of-words. In *Proceedings of the 23rd international conference on Machine learning*, pages 977 – 984, Pittsburgh, Pennsylvania, 2006.

[6] Tanveer A Faruquie, Prem K Kalra, and Subhashis Banerjee. Time based activity inference using latent dirichlet allocation. In *British Machine Vision Conference*, London, UK, 2009.

[7] J. Varadarajan, R. Emonet, and J.-M. Odobez. Probabilistic latent sequential motifs: Discovering temporal activity patterns in video scenes. In *British Machine Vision Conf. (BMVC)*, 2010.