Probabilistic Latent Sequential Motifs: Discovering temporal activity patterns in video scenes

Jagannadan Varadarajan¹² varadarajan.jagannadan@idiap.ch Rémi Emonet¹ remi.emonet@idiap.ch Jean-Marc Odobez¹² odobez@idiap.ch

- ¹ Idiap Research Institute Centre du Parc, Rue Marconi 19, CH-1920 Martigny, Switzerland
- ² École Polytechnique Fédéral de Laussane 1015, Lausanne, Switzerland

Abstract

This paper introduces a novel probabilistic activity modeling approach that mines recurrent sequential patterns from documents given as wordtime occurrences. In this model, documents are represented as a mixture of sequential activity motifs (or topics) and their starting occurrences. The novelties are threefold. First, unlike previous approaches where topics only modeled the co-occurrence of words at a given time instant, our topics model the co-occurrence and temporal order in which the words occur within a temporal window. Second, our model accounts for the important case where activities occur concurrently in the document. And third, our method explicitly models with latent variables the starting time of the activities within the documents, enabling to implicitly align the occurrences of the same pattern during the joint inference of the temporal topics and their starting times. The model and its robustness to the presence of noise have been validated on synthetic data. Its effectiveness is also illustrated in video activity analysis from low-level motion features, where the discovered topics capture frequent patterns that implicitly represent typical trajectories of scene objects.

1 Approach

Among the various methods used for activity analysis from video scenes, topic models like pLSA [4] or LDA [2] have given promising results in discovering dominant activity patterns and abnormality detection through co-occurrence analysis of low-level features [6]. But, one important challenge in these bag-of-words methods is the actual modeling of temporal information: by relying only on the analysis of unordered word cooccurrence within a time window, most topic models fail to represent the sequential nature of activities. Few approaches have been proposed to incorporate temporal information in topic models[1, 5]. However, this was done to represent the dynamics of topic distributions over time. As in videos multiple activities occur simultaneously, applications of these models in video scenes [3] require manually segmented set of clips synchronized with traffic signals to infer temporal information. Considering these issues, our method models each topic as a sequential motif and the starting times of these motifs are directly learnt so as to synchronize multiple occurrences of these temporal motifs.

2 Model

Figure 1 presents the generative model of our approach. Let D be the number of documents d in the corpus, each having N_d words and spanning T_d discrete time steps. Let $V = \{w_i\}_{i=1}^{N_w}$ be the vocabulary of words that can occur at any given instant $t_a = 1, ... T_d$. A document is then described by its count matrix $\mathbf{n}(w, t_a, d)$ indicating the number of times a word w occurs at the absolute time t_a within the document. These documents are generated from a set of N_z topics $\{z_i\}_{i=1}^{N_z}$ assumed to be temporal patterns $p(w, t_r | z)$ with a fixed maximal duration of T_z time steps (i.e. $0 \le t_r < T_z$), where t_r denotes the relative time at which a word occurs within a topic, and that can start at any time instant t_s within the document.

2.1 Generative Process

The actual process of generating all triplets (w,t_a,d) which are counted in the frequency matrix $\mathbf{n}(w,t_a,d)$ works as follows:

- draw a document d with probability p(d);
- draw a latent topic z ~ p(z|d), where p(z|d) denotes the probability that a word in document d originates from topic z;
- draw the starting time $t_s \sim p(t_s|z,d)$, where $p(t_s|z,d)$ denotes the probability that the topic z starts at time t_s within the document d;
- draw a word $w \sim p(w|z)$, where p(w|z) denotes the probability that a particular word w occurs within the topic z;

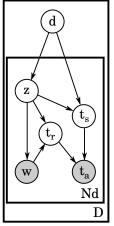


Figure 1: pLSM generative model.

- draw the relative time $t_r \sim p(t_r|w,z)$, where $p(t_r|w,z)$ denotes the probability that the word w within the topic z occurs at time t_r ;
- set $t_a = t_s + t_r$, which assumes that $p(t_a|t_s,t_r) = \delta(t_a (t_s + t_r))$, that is, the probability density function $p(t_a|t_s,t_r)$ is a Dirac function. Alternatively, we could have modeled $p(t_a|t_s,t_r)$ as a noise process specifying uncertainty on the time occurrence of the word.

Ultimately our goal is to discover the topics and their starting times given the set of documents $n(w,t_a,d)$. This is a difficult task since the topic occurrences in the documents overlap temporally. The estimation of the model parameters Θ can be done efficiently using an EM algorithm.

3 Results and conclusion

We demonstrate the strength of our model using a synthetic dataset and two real life traffic datasets. The performance of the algorithm is also shown in the presence of various noise levels which disturb the co-occurrences of words. Furthermore, we also discuss the effect of various model parameters such as topic length, number of topics in learning the topics and start times. When applied to real life data, our results are qualitatively consistent with the activities occurring in the scene. Quantitatively, performance measures also suggest the effectiveness of the method when applied to event detection task. Although the method was demonstrated for activities in a video, we believe that it can have wide applications where sequential patterns need to be extracted.

- [1] D. Blei and J. Lafferty. Dynamic topic models. In *Proceedings of the* 23rd International Conference on Machine Learning, 2006.
- [2] D. M. Blei, A.Y. Ng, and M.I. Jordan. Latent dirichlet allocation. *Machine Learning Research*, (3):993–1022, 2003.
- [3] 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.
- [4] T. Hofmann. Unsupervised learning by probability latent semantic analysis. *Machine Learning*, 42:177–196, 2001.
- [5] Timothy Hospedales, S. Gong, and Tao Xiang. A markov clustering topic model for mining behavior in video. In *ICCV*, Kyoto, Japan, 2009.
- [6] J. Varadarajan and J.M. Odobez. Topic models for scene analysis and abnormality detection. In *ICCV-12th International Workshop on Visual Surveillance*, 2009.