Multimed Tools Appl (2018) 77:10669–10684
DOI 10.1007/s11042-017-4760-4

A dynamic causal topic model for mining activities from complex videos

Yawen Fan[1] · Quan Zhou[1,2] · Wenjing Yue[1] · Weiping Zhu[1,3]

Received: 23 February 2017 / Revised: 18 April 2017 / Accepted: 25 April 2017 / Published online: 10 May 2017

Abstract In this paper, a novel probabilistic topic model is proposed for mining activities from complex video surveillance scenes. In order to handle the temporal nature of the video data, we devise a dynamical causal topic model (DCTM) that can detect the latent topics and causal interactions between them. The model is based on the assumption that all temporal relationships between latent topics at neighboring time steps follow a noisy-OR distribution. And the parameter of the noisy-OR distribution is estimated by a data driven approach based on the idea of nonparametric Granger causality statistic. Furthermore, for convergence analysis during model learning process, the Kullback-Leibler between the prior and the posterior distributions is calculated. At last, using the causality matrix learned by DCTM, the total causal influence of each topic is measured. We evaluate the proposed model through experimentations on several challenging datasets and demonstrate that our model can identify the high influence activity in crowded scenes.

✉ Yawen Fan
ywfan@njupt.edu.cn

Quan Zhou
quan.zhou@njupt.edu.cn

Wenjing Yue
yuewj@njupt.edu.cn

Weiping Zhu
zwp@njupt.edu.cn

1 Key Lab of Ministry of Education for Broad Band Wireless Communication and Sensor Network Technology, Nanjing University of Posts and Telecommunications, Nanjing, China

2 School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China

3 Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada

Keywords Topic models · Video surveillance · Activity analysis

1 Introduction

Probabilistic topic models(PTM) are originally proposed for text processing, such as Probabilistic Latent Semantic Analysis (PLSA) [11], Latent Dirichlet Allocation (LDA) [2], Hierarchical Dirichlet processes(HDP) [23] and their extensions. More recently they have been shown to be successful in video surveillance [13], especially for complicated scenes. The topic model can work directly on low-level motion features, e.g., optical flow. This strategy offers a practical way of bypassing the problems of unreliable detection and tracking in complex scenes. For instance, Wang et al. [25] presented a hierarchical topic model based on optical flow features to detect activities and their interactions in crowded scenes, while Li et al. [16] proposed a hierarchical PLSA model for abnormal behaviors detection by using spatio-temporal features.

Although such approaches offer an advantage in discovering scene activities, most topic models fail to represent the temporal information contained in video streams. Currently, there are two main categories of approaches for including temporal information in the model. The first category is based on word sequences representation [4, 6, 24]. For instance, Varadarajan et al. [24] proposed a novel model to discovering dominant activity motifs from multivariate temporal sequences by introduction of latent variables representing the motif start times. The second category has focused on higher level variables [12, 14, 26]. In these cases, the dynamics of topic distributions over time are modeled. For instance, Hospedales et al. [12] introduced a Markov clustering topic model for scene level behaviors learning, which can dynamically identify irregular spatiotemporal patterns in video. But each activity is still considered as co-occurring of unordered words.

Different from the above mentioned models, we propose a dynamic causal topic model to consider temporal features by introducing a dynamic concentration parameter. This model can mine visual event structures and their temporal relationships simultaneously. The contribution of this paper is threefold:

- We propose a dynamic causal topic model, that not only can detect the activities by capturing co-occurrence of words, but also can obtain the causal interactions between them by utilizing noisy-OR distribution to model the transition of topics over time. For estimating the parameter of the noisy-OR distribution, we adopt a data-driven algorithm based on the idea of nonparametric Granger causality statistic [20].
- For convergence analysis of the model inference process, we develop a new method based on the Kullback-Leibler (KL) divergence between the expected posterior distribution and the expected prior distribution.
- By using the measure of total causal influence, it is possible to identify the high influence topic in a matter of scenes and context.

The rest of this paper is organized as follows. In Section 3.2, the principle of the proposed model is introduced. Section 3.3 gives a detailed description of the parameter estimation. Experimental results are analyzed in Section 4. In this section, the application of the DCTM model to the mining of activities in crowded surveillance videos is explained. We conclude this paper and give future work in Section 5.

2 Related work

Topic models were originally proposed in natural language processing and were designed to identify the semantic structure or 'hidden' structure of a collection of documents containing unordered words. Recently, however, topic models have been frequently used for activity analysis from surveillance video data. In topic model, each document is a collection of words. Activities are seen as a cluster of words. The components of document are activities. For instance, Li et al. [16] proposed a two layer PLSA model for abnormal behaviors detection. Wang et al. [25] presented a hierarchical topic model based on optical flow features to detect activities and their interactions in crowded scenes. Song et al. [22] use a two-staged LDA model to detect both activity and interaction patterns in the video.

However, the above topic model methods mainly focus on spatial co-occurring of the words and may not be appropriate for time series data. To overcome this issue, many solutions have been proposed to include temporal information in the modeling. For instance, Blei and Lafferty [1] propose the Dynamic Topic Model (DTM) for text documents which uses state space models on the natural parameters of the multinomial distributions that represent the evolution of topic word distributions. Different from the DTM, Chua et al. [3] propose a Linear Dynamical Topic Model (LDTM) which utilizes the concepts of Linear Dynamical System to represent the evolution of document topic distributions. In application of activity analysis, temporal extensions of LDA or HDP have also been adapted. For instance, Hospedales et al. [12] extend LDA to model temporal relationships between topics by introducing a Markov chain. Kuettel et al. [14] integrate hidden Markov models with an HDP model to identify multiple temporal topics. These models can only detect temporal interactions at scene level, but the spatial interactions between local activities can't be detected.

Different from the above works, attempts have been made to directly model temporal relationships between topics from the point of causality analysis. McCaffery et al. [19] integrate pLSA with the causal graphical models to learn visual event structures and their temporal relationships simultaneously. But this model structure is simple and is not appropriate for the complex video scenes analysis. In work [5] the HDP is applied to the video data to identify a number of topics of activity in the data. And then the activities are represented as a multivariate point-process. At last the activity temporal dependencies are discovered by the non-parametric Granger causality analysis. In a similar fashion, Kular et al. [15] combines LDA with partial-Granger Causality to discover the main motion and relationships between these motions. In these two works, the causal relationships between activities are layered on top of pre-learned visual topics but they play no role in the mining of the visual topics themselves. To overcome the above limitations, we propose a dynamic causal topic model jointly learning of topics and their causal relationships, which can benefit mutually each other resulting in an optimal solution. And this generative model could be combined with discriminative model, such as SVM [7–10], to realize classification tasks.

3 The proposed framework

In this section, we first give an overview of the model structure and illustrate the model principle for including temporal information contained in videos. Based on model assumptions, a detailed description of the generative process of the model is provided. In the second part, parameter estimation and convergence prediction algorithms are derived, with a special effort on the causation matrix estimation, the KL divergence between expected posterior and prior distributions. Finally the algorithm of causality based topic ranking is proposed.

3.1 Motivation

Complex video surveillance scenarios usually contain multiple activities occurring at specific place and time and many stories chains over time. Therefore, how to use reasonable model structure and parameter learning method to describe the scene is an important and challenging issue. In traditional topic model, there is a hidden structure of data, i.e., the distributions over words and the distributions over topics. For video analysis, words are quantified video features and documents are video clips. Topics are defined as distributions over words, i.e., activities. And each document corresponds a topic distribution.

The subject of this work is to discover activities and the temporal interactions between them. To add the time structure in the topic model, this work introduces the research results of causal analysis related fields. Videos are time-series data. Thus the discovered activities are also the multivariate time series. The causal relationship between the activities can be calculated by the non-parametric Granger method [5]. In the process of topic reasoning, high-level causal feedback is used as the a priori information, that is, the causal relationship between the activities is used to improve the detection performance of the underlying activity. The main contribution of this work is the use of a Granger Causality statistic to explicitly measure the temporal relationships among topics, and code this information into the model's parameter learning process by the Noisy-OR hypothesis.

3.2 Modeling assumptions

Figure 1 shows the probabilistic graphical representation of DCTM using plate diagram. In essence, DCTM is a combination of Latent Dirichlet Allocation (LDA) and the noisy-OR distribution. A temporal correlation measure is needed to compare between the latent topics. While knowing how topics temporally correlate should help the model make better predictions for their future adoption behavior in the next time, the topic distribution at each time step is inferred by the latent topic variable conditioned on the words written in each time step and the topic-item distributions.

The main assumption of the proposed model is that, the topic-item distribution remains static over time, while the topic distribution evolves over time through noisy-OR style distribution conditioned on the inferred latent variables in both previous time step and the current time step. The occurrence of a topic not only depends on the variables in current time step, but also the variables in previous time step. The noisy-OR distribution is a human causal learning standard model. It has a simple and generative causal structure. For these reasons, we use noisy-OR distribution to describe the temporal relationships between latent topics at neighboring time steps. And the notation of the noisy-OR will be denoted by $\mathcal{NOR}(\cdot)$.

The generative process for DCTM is given in the following procedure:

1. For each topic $k \in \{1, ...K\}$, draw a Dirichlet word-topic distribution,

$$\Phi_k \sim Dir(\beta),$$

where β is a V dimensional prior vector of Φ_k.

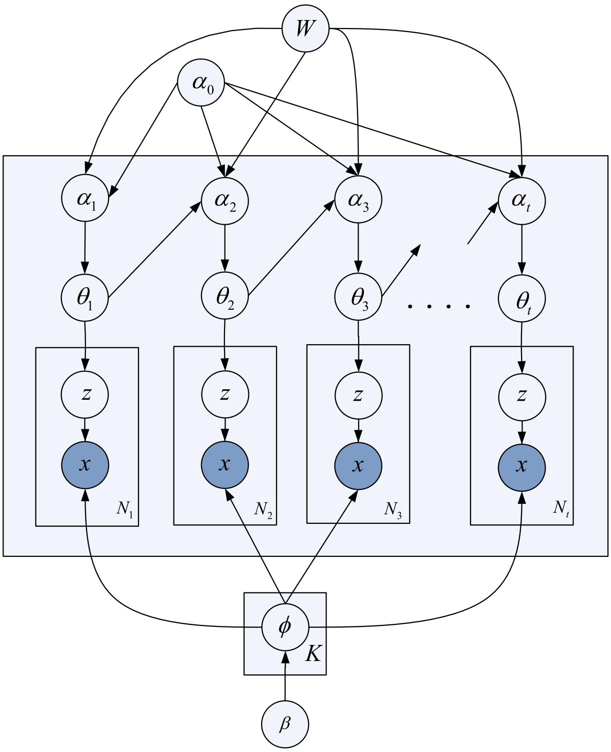


Fig. 1 Graphical representation for the proposed model

2. For each document d_t at the time step t:
   - Generate α_t with the previous time step variables,

$$\alpha_t \propto NOR(\theta_{t-1}, \mathrm{W}) + \alpha_0,$$

     where α_t is a K dimensional prior vector of θ_t ; and $\mathcal{NOR}(\theta_{t-1}, W)$ represents the noisy-OR distribution which depends on the topic distribution at the time step $t-1$ and a causal matrix W.
   - Draw a Dirichlet topic distribution $\theta_t \sim Dir(\alpha_t)$.
3. For each word $i \in \{1, ...N_t\}$ in document d_t:
   - Draw an topic $z_{i,t} \sim Multi(\theta_t)$;
   - Sample a word conditioned on topic variable $x_{i,t} \sim Multi(\Phi_k)$

3.3 Parameter estimation

For topic models, the key inferential problem needed to solve is to compute the posterior distribution of the hidden variables. The approach adopted in this study will be Gibbs sampling algorithm. The causation matrix is estimated by a straight forward approach, directly derived from the data. The Gibbs sampling algorithm is an iterative process. We repeat these steps until a KL Divergence is reached.

3.3.1 Posterior parameters estimation

In this work, we make use of Gibbs sampling algorithm to efficiently obtain an approximation of the posterior distribution of the latent variables. The Gibbs sampler state is the set of the $z_{i,t}$ and its update is derived by integrating out the latent variables θ and ϕ in its conditional probability given the other variables,

$$p(z_{i,t} = k \,\middle|\, Z_{\neg i,t}, D, \alpha_{\mathbf{t}}\ , \beta) \propto \frac{n_{z,x}^{\neg i,t} + \beta_v}{\sum_v (n_{z,v}^{\neg i,t} + \beta_v)} \cdot \frac{n_{t,z}^{\neg i,t} + \alpha_{t,z}}{\sum_k (n_{i,k}^{\neg i,t} + \alpha_{t,k}^{k})}, \tag{1}$$

where D denotes the training documents, and $Z_{\neg i,t}$ denotes all topics excluding $z_{i,t}$; $n_{z,x}$ denotes the counts of word $x_{i,t}$ associated with topic $z_{i,t}$; $n_{t,z}$ denotes the counts of topics $z_{i,t}$ in document d_t.

Based on the sampled latent variables, the posterior distributions are derived,

$$\theta_{t,k} = \frac{n_{t,k} + \alpha_{t,k}}{\sum_k (n_{t,k} + \alpha_{t,k})}, \tag{2}$$

where $n_{t,k}$ denotes the counts of topic k in the document d_t.

$$\phi_{k,v} = \frac{n_{k,v} + \beta_v}{\sum_v (n_{k,v} + \beta_v)}, \tag{3}$$

where $n_{k,v}$ denotes the total counts of word v associated with topic k.

3.3.2 Adaptation of concentration parameter

In LDA, the topic distribution θ is drawn from a fixed hyperparameter α, without considering the temporal information. However, in the streaming data setting, the temporal information is important and should be considered in the generation of θ_t. It is well known that the concentration parameter α has influence on the model. Therefore, we assume that all relationships between topics at the adjacent moment is subject to noisy-OR distribution. As shown in Fig. 2, all topics at current time step will be independent conditioned on the topics at previous time step.

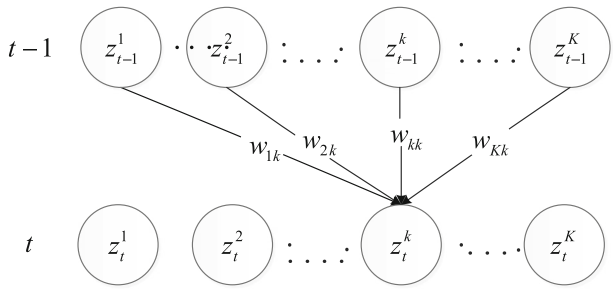


Fig. 2 Temporal relationships between latent topics at neighboring time steps

Given topic mixture proportion θ_{t-1} at the previous time step , the probability that topic k will be present at time t is predicted by the following function,

$$p_t^k = P(z_{t|t-1}^k \,|\theta_{t-1}\,, \mathbf{W}) = 1 - \prod_{m=0}^{K} (1 - \theta_{t-1,m} w_{m,k}), \tag{4}$$

where the matrix $\mathbf{W} = [w_{mk}] \in \mathbf{R}^{K \times K}$ is termed as the causation matrix, essentially quantifies the intensity of all possible causal influences within the system, and w_{mk} denotes the coefficient relating the topics m and k.

To correlate the current parameters α_t with the previous time step variables, α_t is defined as

$$\alpha_t = \mathbf{p}_t + \alpha_0, \tag{5}$$

where $\alpha_0 \in R^K$ represents the initial prior parameter, $\mathbf{p}_t = [p_t^k] \in R^K$ represents the prior topic distribution. This step distinguishes our model from all other topic models.

3.3.3 Causation matrix estimation

In this section, the causation matrix $\mathbf{W}$ is estimated by a straightforward approach, directly derived from the data. Causal relationships among all topics are established via the Nonparametric Granger causality statistic. During the Gibbs sampling process, each visual word is assigned a topic identification. Therefore, a multivariate point-process can be constructed by considering each topic as a point event. The amount of occurrence of topic z^i in the time interval $(t; t + \delta t)$ is then defined as follows:

$$dN_i(t) = N_i(t + \delta t) - N_i(t), \tag{6}$$

where δt denotes the time resolution; $N_i(t)$ is the number of words associated with topic z^i accumulated in the interval $(0; t)$. The mean intensity of the process is defined as $E\{dN_i(t)\} = \lambda_i \delta t$, and the zero-mean process $P_i(t) = N_i(t) - \lambda_i \delta t$ is considered as a point process. Therefore, a K-dimensional multivariate point-process $P(t) = (P_1(t), P_2(t), ..., P_m(t))^T$ is created for representing the dynamics of topics. The causation matrix will be estimated by the multivariate point-process using Nonparametric Granger causality statistic. First, the spectral matrix of the multivariate point-process is computed using multi-taper method. And then based on spectral factorization, Granger causality from topic z^j to z^i at frequency f is obtained by:

$$G_{P_j \to P_i}(f) = \ln \left(\frac{S_{ii}(f)}{S_{ii}(f) - (\Sigma_{jj} - \Sigma_{ij}^2 / \Sigma_{ii}) \left| T_{ij}(f) \right|^2} \right) \,, \tag{7}$$

where $S_{ii}(f)$ represents auto-spectrum, $T(f)$ denotes the transfer function between processes, and Σ denotes the noise process covariance. The causal score between processes $P_j(t)$ and $P_i(t)$ can be obtained by integrating (7) over frequency,

$$\mathbf{W} = \sum_f G_{p_j \to p_i}(f), \forall i \neq j. \tag{8}$$

At last, the causal matrix will be normalized to have a value between 0 and 1.

3.3.4 Convergence analysis

The use of Markov chain Monte Carlo (MCMC) methods has made topic model inference tractable. However, the convergence of this method can be tough to detect. Since the model is updated in every iteration, we propose a new method based upon the divergence between the expected prior and expected posterior distribution and take it as an indication of iteration convergence. Different from the previous works that use topic model on static data, it is more important to distinguish between the posterior and prior distributions in temporal data.

As the posterior topic distribution θ_{t-1} of time step $t-1$ becomes the prior parameters α_t by noisy-OR distribution. Therefore, to justify the convergence, Kullback-Leibler divergence is deduced to measure the distance between the expected posterior topic distribution $\theta_{t|t}$ and expected prior topic distribution $\theta_{t|t-1}$. The KL Divergence between the two distributions is defined as follows,

$$D_{KL}[E(\theta_{t|t}) \,\|\, E(\theta_{t|t-1})]. \tag{9}$$

The expected prior distribution is defined as

$$E(\theta_{t|t-1}) = \frac{\alpha_t}{\sum\limits_k \alpha_{t,k}}, \tag{10}$$

and the expected posterior distribution is defined as:

$$E(\theta_{t|t}) = \frac{\alpha_t + n_t}{\sum\limits_k (\alpha_{t,k} + n_{t,k})}, \tag{11}$$

where $n_{t,k}$ denotes the number of the topic k in the document at time t. Based on the assumption that the performance of the model is improved along with the learning iteration, the KL divergence will be reduced. So the iteration process is repeated until the average KL divergence threshold is reached.

3.3.5 Outline of parameter estimation

Algorithm 1 summarizes the procedure to estimate all parameters of the DCTM model depicted in Fig. 1. It begins by randomly initializing the latent variables, followed by Gibbs sampling iterations which consist of several steps. First, prior parameter is estimated. Then the latent variable is sampled by conditioning on the prior parameters, sufficient statistic from previous iterations and the constant parameters. Based on the sampled latent variables, the posterior parameters are derived. Finally, we estimate the causation matrix. For convergence analysis, the KL between the prior and posterior distributions is calculated. We repeat these steps until a value of KL is reached.

Algorithm 1 Dynamic causal topic model

Input: data stream d_t including totally V unique words
Output: estimated topic-word distribution ϕ, topic distribution θ_t of each document and causation matrix $\mathbf{W}$.
initialize α_0, β, θ , ϕ and $\mathbf{W}$ with the constraint $\sum_{k=1}^{K} \theta_{t,k} = 1, \sum_{v=1}^{V} \phi_{kv} = 1$ and $\mathbf{W} = 0$.
define the total number of topics K
//Gibbs sampling iterations
repeat
  for $t \geq 1$ to the data stream run out **do**
    Estimate concentration parameter α_t by (5);
    Sample topic by (1);
    Update posterior distribution θ_t by (2);
  end for
  Update posterior distribution ϕ by (3);
  Update causality matrix $\mathbf{W}$ by (8);
  Calculate average KL divergence by (9);
until Convergence

3.4 Causality based topic ranking

In the above definition, the ith row in the causation matrix $\mathbf{W}$ is comprised of the causal influences of the corresponding topic i on the other topics. Therefore, the total causal influence of the topic i is measured as the l_1-norm of the ith row in the causation matrix $\mathbf{W}$,

$$T_i = \sum_{j=1}^{K} |w_{ij}| = \sum_{j=1}^{K} w_{ij}. \tag{12}$$

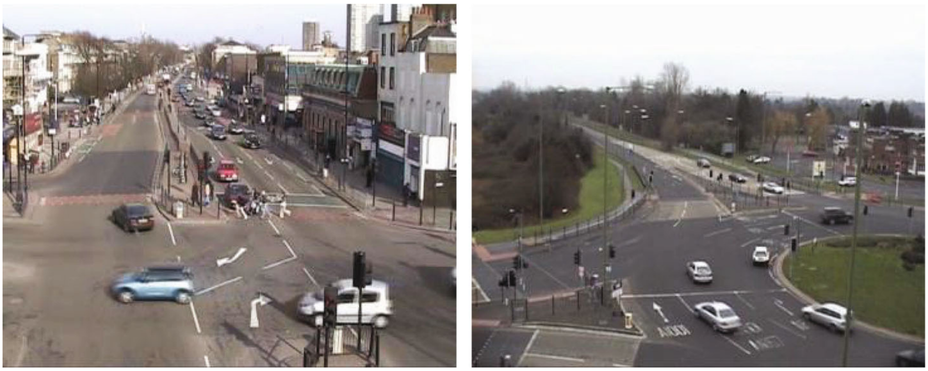
This quantifies the single topic's total causal influences within the system. It may contribute to identify the high influence topic in crowded scenes. Based on the measured causal influence $\{\mathrm{T}_i\}_i^K$, the topic can be ranked in the form,

$$Lowest\ T_{i1} \leq T_{i2} \leq ... \leq T_{iK}\ Hightest. \tag{13}$$

4 Experiments

4.1 Datasets and setting

The experiments were performed on Intel Core i7-2600 machine with 3.4GHz CPU and 4G RAM, and implemented in hybrid programming with C and Matlab. Experimental results are demonstrated on two traffic video sequences selected from the QMUL dataset at 25fps with a resolution 360×288. As shown in Fig. 3, both two scenes are junctions with traffic lights. Thus, the sequence of activities exhibits spatial-temporal periodicity. Typically, there



(a) intersection dataset (b) roundabout dataset

Fig. 3 Dataset

are several flows at a time, and each flow may last for a period. We assess the proposed method with both qualitative and quantitative evaluations.

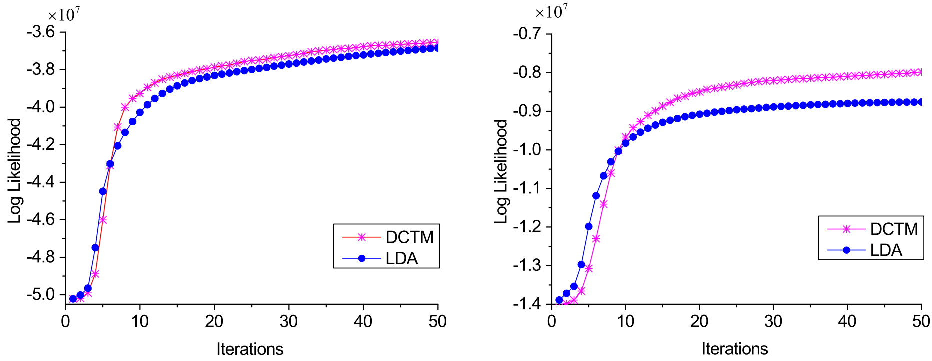
In this study, low-level motion feature is adopted to represent video. To apply the topic model on video data, we need to establish a bag-of-words representation. Firstly, we spatially divide the scene into a 36×29 grid with a spacing of 10 pixels, and then, the motion direction is quantized into 8 orientations for each cell. Therefore, a codebook with $36 \times 29 \times 8$ visual words is constructed. For the input video sequences, they are temporally segmented into 3-s-long non-overlapping video clips. TV-L1 algorithm [27] is used to extract optical flow vectors from each frame. Optical flow vectors with the magnitudes less than 0.8 are ignored. By spatial and directional quantization, the denoised flow vectors are indexed to one of the visual words in the codebook. The visual document is composed by the words accumulated over the corresponding video clip. The dataset sizes are detailed in Table 1. The proposed model is an extension of LDA, so it is used as the baseline model. For both models, we set 22 topics for the intersection scene, and set 26 topics for the roundabout scene. And set $\alpha_0 = 0.5$, $\beta = 0.04$.

4.2 Convergence of log likelihood

We evaluate the convergence of the log likelihood for the proposed DCTM and the LDA model. Figure 4a and b show how the log likelihood varies with the number of iterations for both sequences. Due to the intractable of sum over correlated z, the importance sampler is adopted to approximate the likelihood. As can be seen from Fig. 4, DCTM model achieves the higher likelihood compared with LDA. This indicates that the proposed model can better model the activities contained in the test video. It is obvious that the high-level causal feedback can improve the low level topic prediction.

Table 1 Dataset sizes

Dataset	Documents	Topics	Words	Period(min)
intersection	1199	21	6821431	60
roundabout	1246	26	2136908	62.3

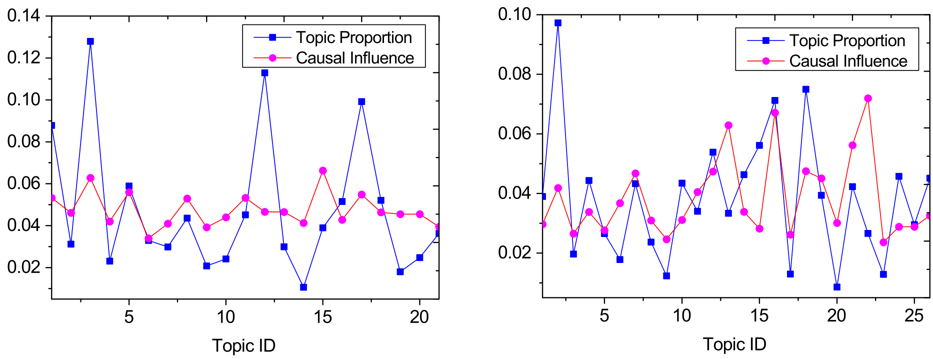


(a) intersection scene (b) roundabout scene

Fig. 4 Log likelihood versus the number of iterations

Fig. 5 Top 4 topics for intersection scene

Fig. 6 Top 4 topics for roundabout scene



(a) intersection scene (b) roundabout scene

Fig. 7 Total mixture proportion and total causal influence for topics

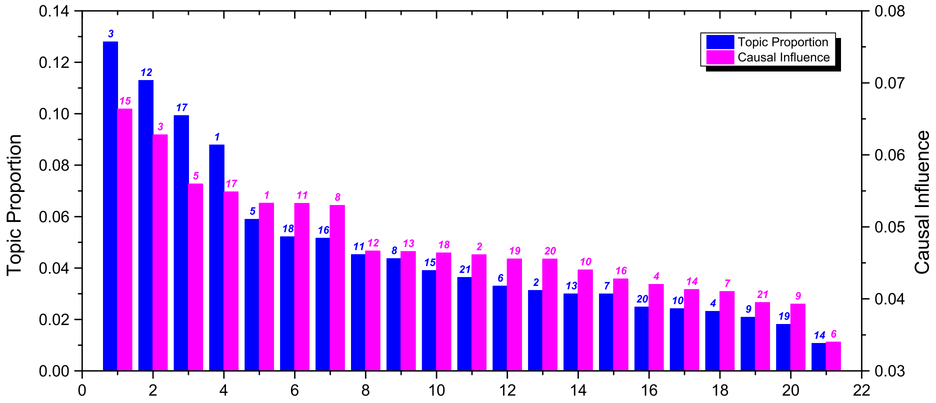


Fig. 8 Total mixture proportion and total causal influence based topic ranking for the intersection scene

4.3 Causality based topic ranking

For the intersection scene, the motion distributions of the top four topics are selected as shown in Fig. 5. Topics 15 and 3 indicate the horizontal traffic flow with different directions, respectively. Topic 5 indicates upward vehicles traffic flow. Topic 8 indicates right-turning traffic flow. For the roundabout scene, the motion distributions of the top four topics are also selected as shown in Fig. 6. Topics 22 and 16 indicate left to right traffic flows at different zones. Topic 13 indicates right-turning traffic flow. Topics 21 indicates left-turning traffic flow in the far field.

Compared with the LDA model, the structure of the proposed DCTM is complicated with more parameters. So it could provide more semantic information. Based on the causation matrix **W**, the total causal influence of each activity is computed. For comparison, the total mixture proportions(TMP) of topics are given by,

$$Prop(z^k) = \frac{\sum_t n_{t,k}}{\sum_t \sum_k n_{t,k}}, \tag{14}$$

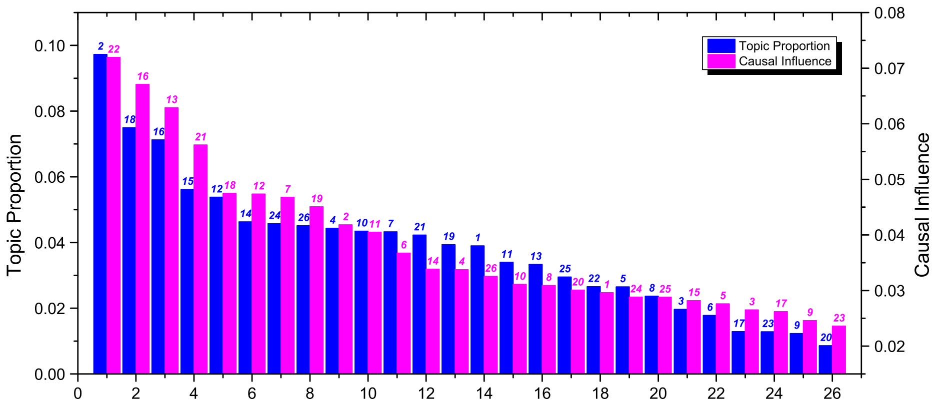


Fig. 9 Total mixture proportion and total causal influence based topic ranking for the roundabout scene

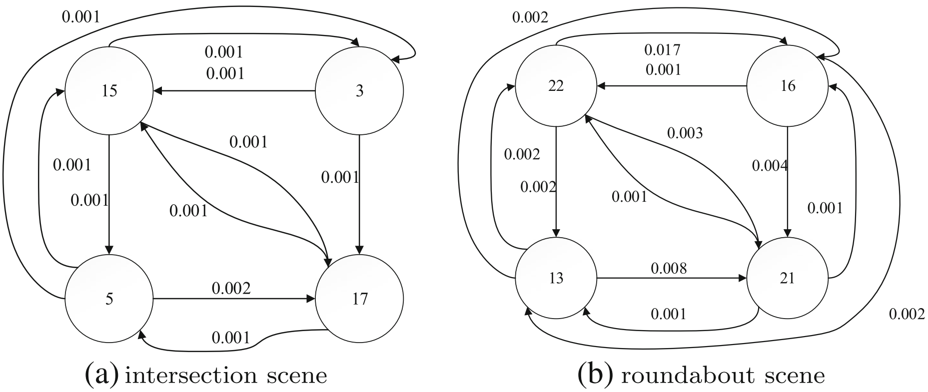


Fig. 10 Causal graph

where $n_{t,k}$ denotes the counts of the topic k in the document at time t, which can be obtained from the sufficient statistics. As shown in Fig. 7, the total mixture proportions and total causal influence over topics are represented by curves. It is obviously that there are some differences in trend in the two series.

To provide further insights on the difference, topics are ranked based on the TMP and TCI respectively. As can be seen from Figs. 8 and 9, the ordered topics are represented by bars, where the labels on top of the bars denote topic IDs. A large number of topic mixture proportions does not always mean high influence. This is in line with common sense. Therefore TCI-based topic ordering is more helpful for video surveillance. The video surveillance system could pay more attention on the high-impact activity to avoid some anomalies.

To demonstrate the interaction between activities, the causality matrix is represented by a directed graph, where nodes denote topics and edges denote the causal relations between them. Figure 10 shows the causal graphs corresponding with the top 4 activities.

5 Conclusions

Activity analysis for complex scene is a challenging problem, especially in situations where multiple activities are occurring simultaneously. In this study, a novel topic model is proposed to mine activities and their causal influence in the complex scenes. The concentration parameter of the Dirichlet distribution is updated at every time step through the noisy-OR distribution. To estimate the causal parameter of the noisy-OR distribution, the topics are deemed as multivariate point process. Then the causal parameter is obtained by performing the non-parameter Granger causal analysis on the pairs of point process. Additionally, we can take usage of the total causal influence to identify the high influence activity. This is useful for video surveillance.

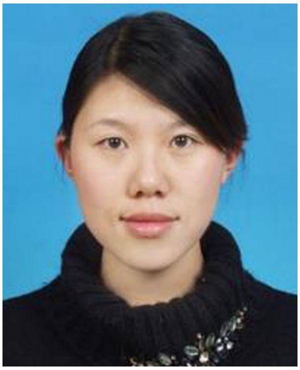
In future studies, we will carry out more experiments on different sequences to evaluate the performance of the proposed model. Moreover, causal grouping algorithms will be studied to mine the temporal rules among the activities. And how to use this model to identify the leader in the crowd based on the trajectory features is also a question that will be considered and investigated. Except for behavioral analysis, the promising applications of this model include under water image classification [17, 18], video streaming [28] and social event analysis [21] etc.

Acknowledgements The authors would like to thank the associated editor and all the anonymous reviewers for their valuable comments and suggestions. This work was partly supported by Natural Science Foundation of Jiangsu Province(Grant No. BK20160908), and the Key Lab of Broadband Wireless Communication and Sensor Network Technology(Grant No.NYKL2015012), and the National Natural Science Foundation of China (Grant No. 61401228, 61501253) and the Basic Research Program of Jiangsu Province(Natural Science Foundation)(Grant No. BK20151506), and the China Postdoctoral Science Foundation(Grant No. 2015M581841), and the Postdoctoral Science Foundation of Jiangsu Province (Grant No. 1501019A), and the Priority Academic Program Development of Jiangsu Higer Education Institutions(PAPD), and the Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology(CICAEET), and the Nanjing University of Information Science and Technology Research Foundation for Talented Scholars (Grant No. 2015r014).

References

1. Blei DM, Lafferty JD (2006) Dynamic topic models[C]. In: Proceedings of the 23rd international conference on machine learning. ACM, pp 113–120
2. Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation[J]. J Mach Learn Res 3:993–1022
3. Chua FCT, Oentaryo RJ, Lim EP (2015) Using linear dynamical topic model for inferring temporal social correlation in latent Space[J]. Computer Science
4. Emonet R, Varadarajan J, Odobez J (2011) Extracting and locating temporal motifs in video scenes using a hierarchical non parametric Bayesian model[C]. IEEE Computer Vision and Pattern Recognition 3233–3240
5. Fan Y, Yang H, Zheng S et al (2013) Video sensor-based complex scene analysis with Granger causality[J]. Sensors 13(10):13685–13707
6. Faruquie TA, Kalra PK, Banerjee S (2009) Time based activity inference using latent dirichlet allocation[C]. BMVC 1–10
7. Gu B, Sheng VS (2016) A robust regularization path algorithm for v-support vector classification[J]. IEEE Transactions on Neural Networks and Learning Systems 28(5):1241-1248
8. Gu B, Sun X, Sheng VS (2016) Structural minimax probability machine. IEEE Transactions on Neural Networks and Learning Systems. doi:10.1109/TNNLS.2016.2544779
9. Gu B, Sheng VS, Tay KY et al (2015) Incremental support vector learning for ordinal regression[J]. IEEE Transactions on Neural Networks and Learning Systems 26(7):1403–1416
10. Gu B, Sheng VS, Wang Z et al (2015) Incremental learning for v-support vector regression[J]. Neural Netw 67:140–150
11. Hofmann T (2001) Unsupervised learning by probabilistic latent semantic analysis[J]. Mach Learn 42(1-2):177–196
12. Hospedales T, Gong S, Xiang T (2009) Markov clustering topic model for mining behaviour in video[C]. In: IEEE international conference on computer vision, IEEE, pp 1165–1172
13. Kinoshita A, Takasu A, Adachi J (2015) Real-time traffic incident detection using a probabilistic topic model[J]. Inf Syst 54(C):169–188
14. Kuettel D, Breitenstein MD, Gool LV et al (2010) What's going on? Discovering spatio-temporal dependencies in dynamic scenes[C]. IEEE Computer Vision and Pattern Recognition 1951–1958
15. Kular D, Ribeiro E (2015) Analyzing activities in videos using latent dirichlet allocation and granger causality[C]. In: International symposium on visual computing. Springer International Publishing, pp 647–656
16. Li J, Gong S, Xiang T (2008) Global behaviour inference using probabilistic latent semantic analysis[C]. BMVC 3231:3232
17. Li Y, Lu H, Li J et al (2016) Underwater image de-scattering and classification by deep neural network[J]. Comput Electr Eng 54:68–77
18. Lu H, Li Y, Nakashima S et al (2016) Single image dehazing through improved atmospheric light estimation[J]. Multimedia Tools and Applications 75(24):17081–17096
19. Mccaffery JP, Maida AS (2013) Toward a causal topic model for video scene analysis[C]. In: International joint conference on neural networks, pp 1–8
20. Nedungadi A, Rangarajan G, Jain N, Ding M (2008) Analyzing multiple spike trains with nonparametric granger causality. J Comput Neurosci 27:55–64
21. Qian S, Zhang T, Xu C et al (2016) Multi-modal event topic model for social event analysis[J]. IEEE Trans Multimedia 18(2):233–246
22. Song L, Jiang F, Shi Z et al (2011) Understanding dynamic scenes by hierarchical motion pattern mining[C]. In: IEEE international conference on multimedia and expo (ICME), 2011. IEEE, pp 1–6

23. Teh YW, Jordan MI, Beal MJ, Blei DM (2006) Hierarchical dirichlet processes. J Am Stat Assoc 101(476)
24. Varadarajan J, Emonet R, Odobez JM (2013) A sequential topic model for mining recurrent activities from long term video Logs[J]. Int J Comput Vis 103(1):100–126
25. Wang X, Ma X, Grimson WEL (2009) Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models[J]. IEEE Trans Pattern Anal Mach Intell 31(3):539–555
26. Xue J, Eguchi K (2016) Sequential correspondence hierarchical dirichlet processes for video data analysis[C]. ACM ACM 229–233
27. Zach C, Pock T, Bischof H (2007) A duality based approach for realtime TV-l, 1 optical Flow[J]. Lect Notes Comput Sci 4713(5):214–223
28. Zhou L, Hu RQ, Qian Y et al (2013) Energy-spectrum efficiency tradeoff for video streaming over mobile ad hoc networks[J]. IEEE J Sel Areas Commun 31(5):981–991



Yawen Fan She received BE and MS degrees in electronic engineering from Hohai University, Nanjing, China in 2003 and 2006, respectively. She has received the doctor degree in EE department from Shanghai Jiao Tong University, Shanghai, China. She is now an assistant professor at Nanjing University of Posts and Telecommunications, Nanjing, P. R. China. Her research interests are intelligent video surveillance and video analysis and understanding.

Quan Zhou He received the M.S. degree and Ph.D. degree in communication and information system in 2006 and 2013, respectively, from Huazhong University of Science and Technology, China. He is now an associate professor of Nanjing University of Posts and Telecommunications. His research interests include computer vision and pattern recognition. He has published over 20 research papers in SCI journals (e.g., IEEE Transactions on Image Processing, IEEE Transactions on Multimedia, Pattern Recognition) and conference (ICIP, ICASSP, ACCV, and ICPR) in image processing and computer vision. He now serves as TPC member or chair of many international conferences and reviewer for a series of SCI journals, including IEEE Transactions on Image Processing, IEEE Transactions on Multimedia, IEEE Transactions on Circuits System for Video Technology, Pattern Recognition, and Neurocomputing. He is member of IEEE.

Wenjing Yue She received his B.Sc. and M.Sc. degrees both from Taiyuan University of Technology in 2003 and 2006 respectively, and received his Ph.D. degree from Shanghai Jiao Tong University in 2010. Now she is an associate professor in Nanjing University of Posts and Telecommunications. Her research interest includes wireless communications and networks, sensor networks.

Weiping Zhu He received the B.E. and M.E. degrees from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1982 and 1985, respectively, and the Ph.D. degree from Southeast University, Nanjing, in 1991, all in electrical engineering. From 1991 to 1992 and from 1996 to 1998, he was a Post-Doctoral Fellow and a Research Associate, respectively, with the Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada. From 1993 to 1996, he was an Associate Professor with the Department of Information Engineering, Nanjing University of Posts and Telecommunications. From 1998 to 2001, he was with high-tech companies in Ottawa, ON, Canada, including Nortel Networks and SR Telecom Inc. Since 2001, he has been a full-time faculty member with the Department of Electrical and Computer Engineering, Concordia University, where he is currently a Full Professor. Since 2008, he has been also an Adjunct Professor with the Nanjing University of Posts and Telecommunications. His research interests include digital signal processing fundamentals, speech and audio processing. He served as an Associate Editor of the IEEE Transactions on Circuits and Systems–I and–II. He was also a Guest Editor of the IEEE Journal on Selected Areas in Communications. He was the Secretary of Digital Signal Processing Technical Committee (DSPTC) of the IEEE Circuits and System Society during June 2012-May 2014, and the Chair of the DSPTC during June 2014–May 2016.