



Efficient Analysis of Traffic Intersection Scenes by Employing Traffic Phase Information

P. Ahmadi^{*(C.A.)} and I. Gholampour^{**}

Abstract: Analyzing motion patterns in traffic videos can be employed directly to generate high-level descriptions of their content. For traffic videos captured from intersections, usually, we can easily provide additional information about traffic phases. Such information can be obtained directly from the traffic lights or through traffic lights controllers. In this paper, we focus on incorporating additional information to analyze the traffic videos more efficiently. Using side information on traffic phases, the semantic of motion patterns from traffic intersection scenes can be learned more effectively. The learning is performed based on optical flow features extracted from training video clips, and applying them to supervised topic models such as MedLDA and MedSTC. Based on such models, any video clip can be represented based on the learned patterns. Such representations can be further exploited in scene analysis, rule mining, abnormal event detection, etc. Our experiments show that employing side information in intersection video analysis leads to improvement in discovering scene pattern. Moreover, supervised topic models achieve about 4% improvement in abnormal event detection, compared to the unsupervised ones, in terms of area under ROC.

Keywords: Motion Patterns, Traffic Intersection, Traffic Phase, Supervised Topic Models.

1 Introduction

COMMONLY, control rooms that are observing places such as crowded streets, are witness of fully packed and motioned scenes. Among these, there may be repetitive behaviors in some scenes that follow certain patterns, known as motion patterns. There is great tendency to go over the motion patterns so as to obtain highly accurate scene interpretation with detailed contents. Identification of these motion patterns results in a semantic scene model that makes it easier to analyze traffic scenes [1]. But, it may also pose some difficulties since most of the current models fail to properly analyze these scenes [1]. Previous models of crowded street video content analysis [2, 3] were built

upon trajectory data. These models require finding precise location of street objects and tracking them which seems to be a difficult task because there are some environmental factors (noise, light, meteorological conditions, traffic jam, etc.) that disturb the analysis [4].

To bring the issue under control, there have been some attempts [1, 5-10] so as to discover more apparent behavior pattern of traffic scenes by employing motion features like optical flows and proposing more precise models, known as topic models. To do so, probabilistic topic models were presented, such as probabilistic Latent Semantic Analysis (pLSA) [11] and Latent Dirichlet Allocation (LDA) [12], to identify hidden topics of a whole scene in a more precise way. A non-probabilistic topic model proposed by Zhu and Xing (2011) [13], Sparse Topical Coding (STC), which can detect hidden depiction of a huge set of data.

Using topic models, e.g. LDA and HDP, researchers [5, 6] managed to analyze the behavior pattern of motions in traffic scenes. Song *et al.* [7] trained a two-level Latent Dirichlet Allocation model both for single- and multi-agent activity patterns. Li *et al.* [8] produced a two-level cascaded Latent Dirichlet Allocation model to study both local and global features

Iranian Journal of Electrical and Electronic Engineering, 2019.

Paper first received 15 April 2017 and accepted 24 June 2018.

* The author is with the IT Research Faculty, Iran Telecommunication Research Center, Tehran, Iran.

E-mail: p.ahmadi@itrc.ac.ir

** The author is with the Electronics Research Institute, Sharif University of Technology, Tehran, Iran.

E-mail: jmangh@sharif.edu

Corresponding Author: P. Ahmadi.

of the content in two levels respectively, yet the second level lays on the output of the first level. Emonet *et al.* in [9] proposed a sequential topic model in order to examine repetitive pattern of behaviors in traffic videos. Rana *et al.* [10] used a fast and first rank Robust Principal Component Analysis (RPCA) to analyze foreground contents (e.g. pixels), that are the input of DPMM, so as to facilitate gradual understanding of the content. Other researchers [1] have also used STC to gain a properly well learning of topic models.

Usually, in traffic intersections, we can easily have access to the information about traffic phases. Such information can be obtained directly from the traffic lights or through traffic light controllers. Such controllers are provided in various types, including fixed time, programmable pre-time, and intelligent controllers. As an example, SCATS (Sydney Coordinated Adaptive Traffic System) is an intelligent traffic management system which has been used in various countries since 1982. In this paper, we focus on employing such additional information to analyze the traffic videos more efficiently. Using auxiliary information on traffic phases, semantic motion patterns can be learned from a traffic intersection scene. The learning is done by applying low-level motion features to supervised topic models such as MedLDA [14] and MedSTC [13]. The learned motion patterns can further be employed in scene analysis applications, such as rule mining, abnormal event detection, etc. The remainder of the paper is organized as follows: Section 2 is devoted to theoretical background of supervised topic models to support the next sections. Section 3 explains our proposed method for video representation, learning motion patterns and abnormality detection. Section 4 shows our experimental results. Section 5 concludes the paper.

2 Theoretical Background

A review of previously studied frameworks on supervised topic models is given below.

2.1 MedLDA

Latent Dirichlet Allocation (LDA) is a widely known topic model since it primarily applies Dirichlet on word and topic distribution. LDA was successful in enhancing the level of pervious models' performance (e.g. pLSA).

In the whole body of D documents, each document is a combination of K topics. Each topic k is a polynomial distribution of words indicated with $\beta = \{\beta_k\}$. In the given documents, α is a K -dimensional Dirichlet prior. K -dimensional parameter θ_d of polynomial distribution is derived from Dirichlet distribution $Dir(\theta_d, \alpha)$ for every given document d . For every given word w_{dn} of document d , a topic z_{dn} is derived with probability θ_{dk} , and the word w_{dn} is also derived from the polynomial distribution indicated by $\beta(z_{dn})$. α and β are the hyper-parameter of a prior distribution that are to be modified.

If we assume that α and β are given, for topic mixture of a joint distribution θ , a group of N topics z , and a group of N words w , we have [12]:

$$p(\theta_d, z_d, w_d | \alpha, \beta) = p(\theta_d | \alpha) \prod_{n=1}^N p(z_{dn} | \theta_d) p(w_{dn} | z_{dn}, \beta) \quad (1)$$

To express the inference accurately, the integrated likelihood $P(w_n | \alpha, \beta)$ and the resulting posterior distribution $P(\theta_d, z_d | \alpha, \beta)$ are troublesome. So, a statistical inference method (i.e. variational Bayesian approximation) is required to estimate $P(\theta_d, z_d | \alpha, \beta)$.

For the classification problem, a Supervised Topic Model of LDA (MedLDA) is presented in [14] and depicted in Fig. 1.

MedLDA attempts to discover distinct topics that are more suitable for classification of texts and documents [14]. By having a large-margin classifier η , for a document of class y_d , joint distribution of hidden and visible variables of MedLDA can be obtained as follows [14]:

$$p(\theta_d, z_d, w_d, y_d | \alpha, \beta, \eta) = p(\theta_d | \alpha) \prod_{n=1}^N p(z_{dn} | \theta_d) p(w_{dn} | z_{dn}, \beta) p(y_d | z_d, \eta) \quad (2)$$

Assume a class of $y = \{1, \dots, M\}$ as response variable of classification. Also, assume that a linear discriminant function, $F(y, z_{1:N}, \eta) = \eta_y^T \bar{z}$, in which $\bar{z} = 1/N \sum_n z_n$, η_y is a parameter vector with K dimension and specified class y , and η is a vector with MK dimensions which stacks the elements of η_y . Similarly, F may be defined as $F(y, z_{1:N}, \eta) = \eta^T f(y, \bar{z})$; here, $f(y, \bar{z})$ is a vector of numerical features in which all components are zero except $(y-1)K+1$ to yK that are derived from \bar{z} . A support vector machine is required to construct a decision rule for every single F . Here, a general $q(\eta)$ distribution is examined rather than studying parametric point estimation of η and the average of all possible models and hidden topics is considered in predicting the rule [14]:

$$y^* = \arg \max_y E[\eta^T f(y, \bar{z}) | \alpha, \beta] \quad (3)$$

The integrated latent model for topic discovery and multi-class classification model are defined as follows [14]:

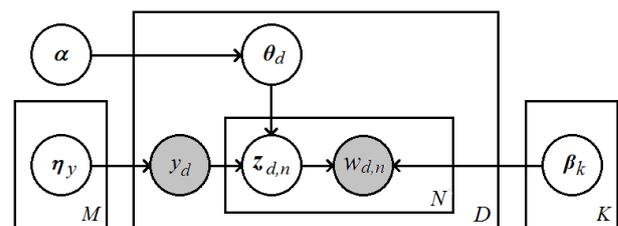


Fig. 1 MedLDA [15].

$$\begin{aligned} \min_{q, \eta, \alpha, \beta, \xi} & L(q) + KL(q(\eta) \| p_0(\eta)) + C \sum_{d=1}^D \xi_d \\ \text{s.t. } & \forall d, y \neq y_d : E[\eta^T \Delta f_d(y)] \geq 1 - \xi_d; \xi_d \geq 0 \end{aligned} \quad (4)$$

where, η is sampled from a prior $p_0(\eta)$, $q(\theta, z|\gamma, \phi)$ is a variational distribution, $L(q) = -E[\log p(\theta, z, w | \alpha, \beta)] - H(q(\theta, z))$ is upper variational bound of $-\log p(w | \alpha, \beta)$, $H(q)$ is the entropy of q ; $\Delta f_d(y) = f(y_d, \bar{z}_d) - f(y, \bar{z}_d)$, and ξ indicates slack variable that absorb errors during data instruction. KL -term is regularizer of the distribution $q(\eta)$. $E[\eta^T \Delta f_d(y)]$ is the predicted margin by which the correct label y_d is preferred to a prediction y .

MedLDA has the potential of finding hidden topic representation $q(\theta, z|\gamma, \phi)$ and the parameter distribution $q(\eta)$ which both anticipates the training data as precisely as possible, and describes the data properly well.

2.1 MedSTC

STC represents a non-probabilistic formulation for topic model and assigns a sparse topic number to every document. Assume that $V = \{1, \dots, N\}$ is a N word vocabulary. $w = (w_1, \dots, w_N)^T$ describes the document as a vector in which I is the index of set of presented words and w_n ($n \in I$) indicates the number presented word n in this document. Assume that $\beta \in \mathbb{R}^{K \times N}$ is a dictionary with K bases, where each row β_k is supposed to be a base for a topic i.e., an n -gram distribution over V . If P is a $(N-1)$ -simplex, then $\beta_k \in P$. STC depicts w as an input in a semantic space that is spanned by a group of automatically lettered bases β and reaches to a highly accurate appearance of the whole body of document. $\theta_d \in \mathbb{R}^K$ indicates document code for document d and $s_{dn} \in \mathbb{R}^K$ indicates word code for word n . When STC minimizes the objective function, denoted by $f(\theta, s, \beta)$, it takes the following form [13]:

$$\begin{aligned} \min_{\theta, s, \beta} & \sum_{d,n} \log p(w_n | s_{dn}, \beta) \\ & + \sum_{d,n} (\gamma \|s_{dn} - \theta_d\|_2^2 + \rho \|s_{dn}\|_1) + \lambda \sum_d \|\theta_d\|_1 \\ \text{s.t. } & \theta_d \geq 0, \forall d \\ & s_{dn} \geq 0 \forall n, d \in I_d \\ & \beta_k \in P \forall k \end{aligned} \quad (5)$$

where, λ , γ , and ρ are nonnegative hyper-parameters defined by users [13] and:

$$p(w_n | s_{dn}, \beta) = \frac{(s_{dn}^T \beta_n)^{w_n} \exp\{-s_{dn}^T \beta_n\}}{w_n!} \quad (6)$$

In [13], a modified version of STC, MedSTC, is presented for classification problems, as illustrated in Fig. 2.

By using surrounding information of document labels, MedSTC can predict representations and get supervised

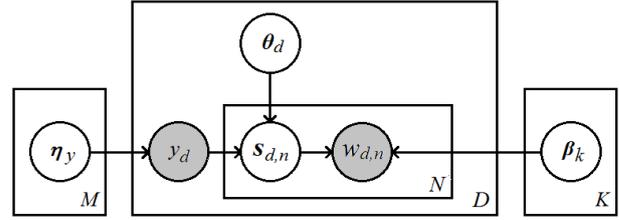


Fig. 2 MedSTC*.

* Ref. [13] does not have this plate diagram; rather, we create this illustration from the description of their model.

dictionary word-perfect [13]. As non-probabilistic STC may be joint with any convex loss function by default, the principle of large margin is developed to specify a classifier that can be favorably explored in MedLDA [14], as well.

Document code θ is particularly considered as an input parameter for a SVM multi-class classifier and it can explain linear discriminant function of $F(y, \theta) = \eta_y^T \theta$, in which $\eta_y \in \mathbb{R}^K$. Assume that η denotes a class of η_y , and $\Delta l(y, y')$ represents a cost function that measures difference between predicted y' and true label y . By assuming $D = \{(w_d, y_d)\}_{d=1}^D$ as a training set, the multi-class loss function is [13]:

$$R_h(\{\theta_d\}, \eta) = \frac{1}{D} \sum_d \max_y [\Delta l(y_d, y) + F(y, \theta_d) - F(y_d, \theta_d)] \quad (7)$$

The max-margin supervised STC (MedSTC) can be described as joint learning a large margin classifier η , getting over a dictionary β , and discovering hidden representations of s and θ . Joint optimization equation can be defined as [13]:

$$\begin{aligned} \min_{\theta, s, \beta, \eta} & f(\theta, s, \beta) + CR_h(\{\theta_d\}, \eta) + \frac{1}{2} \|\eta\|_2^2 \\ \text{s.t. } & \theta_d \geq 0, \forall d \\ & s_{dn} \geq 0, \forall d, n \in I_d \\ & \beta_k \in P, \forall k \end{aligned} \quad (8)$$

Here, C is a positive constant. As regards, document code θ bridges the internal hidden representations to the external supervision.

3 Proposed Method

In this section, we explain our proposed approach for traffic video representation and then learning motion pattern by employing auxiliary information about traffic phases. At the next step, we introduce our method for abnormality detection in traffic videos. The proposed method is based on supervised topic models like MedLDA and MedSTC. Such models have been originally proposed for the purpose of document classification. They utilize the known class labels of training documents to find the unknown labels of test documents. In our proposed method, we assume that the

traffic phase labels of both training and test video clips are available. Based on this assumption, we employ the supervised topic models for more efficient analysis of the traffic scenes.

3.1 Video Representation

After capturing a video, in the first place, it is divided to a series of clips that do not have common scenes. Each clip is regarded as a document.

In order to identify key points and employ them to derive optical flow vectors through Lucas–Kanade method from pairs of succeeding frames, Shi-Tomasi corner detector is used here. To reduce the noise volume, a limit is defined for optical flow vectors' amplitude. In order to develop the vocabulary, optical flow vectors are converted to discrete visible words. Optical flow vectors are indicated by (x, y, α) . The positions of (x, y) are limited to the nearest region on a grid spaced every 10 pixels and the angles of flow vectors, α , are divided into 8 directions. At last, a specified vocabulary is obtained, i.e. $\mathbf{v} = \{v_1, \dots, v_N\}$ with a total of N flow words, where every word contains some data about location and motion direction.

For MedLDA, the video clip can be represented as a vector $\mathbf{w} = \{w_1, \dots, w_N\}$, in which w_n denotes the word n in the clip. For MedSTC, in each clip, a group of flow words gather together over the video frames. After that, the video clip can be represented as a vector $\mathbf{w} = \{w_1, \dots, w_I\}$, in which I is a set of word indexes and w_n denotes the number of appearance of word n in the clip. By employing a topic model, optical flow words that frequently co-occur in a video clip create a motion pattern, which indicates distribution of the vocabulary \mathbf{v} .

3.2 Learning Motion Patterns

Let us assume that we know the traffic phase of each training video clip. For employing the additional information of traffic phases, we propose to utilize the supervised topic models such as MedLDA and MedSTC as the topic model. The traffic phase of each video clip defines a label for the class that the clip belongs to. We use such known labels of training video clips as the variable y (label y_d for the clip d) in the supervised topic models. Then, we can find the topics or motion patterns β by solving the optimization problem (9) using MedLDA.

$$\begin{aligned} \min_{q, \alpha, \beta, \xi} & -E[\log p(\theta, z, w | \alpha, \beta)] - H(q(\theta, z)) \\ & + KL(q(\eta) || p_0(\eta)) + C \sum_{d=1}^D \xi_d \\ \text{s.t.} & E[\eta^T \mathbf{f}(y_d, \bar{z}_d) - \eta^T \mathbf{f}(y, \bar{z}_d)] \geq 1 - \xi_d; \xi_d \geq 0 \\ & \forall d, y \neq y_d \end{aligned} \quad (9)$$

where α is a Dirichlet prior, η is a large-margin classifier, θ is the topic mixture and z is the topic assigned to each video clip. All parameters are as

explained in subsection 2.1.

Similarly, using the known labels of training video clips y in MedSTC, we can find the dictionary of topics or motion patterns β by solving the optimization problem (10).

$$\begin{aligned} \min_{\theta, s, \beta, \eta} & \sum_{d,n} (w_{dn} \log(s_{dn}^T \beta_n) - s_{dn}^T \beta_n) \\ & + \sum_{d,n} (\gamma \|s_{dn} - \theta_d\|_2^2 + \rho \|s_{dn}\|_1) + \lambda \sum_d \|\theta_d\|_1 \\ & + \frac{C}{D} \sum_d \max_y [\Delta I(y_d, y) + \eta_y^T \theta_d - \eta_{y_d}^T \theta_d] \\ & + \frac{1}{2} \|\eta\|_2^2 \\ \text{s.t.} & \theta_d \geq 0, \forall d \\ & s_{dn} \geq 0, \forall d, n \in I_d; \\ & \beta_k \in P, \forall k \end{aligned} \quad (10)$$

where η is a large-margin classifier, θ is the document code and s is the word codes. All parameters are as explained in Subsection 2.2.

The discovered motion patterns can be further employed for finding the motion patterns that occur in a test video clip correspond to a given traffic phase.

3.3 Abnormality Detection

In the training process, the motion patterns β and the large-margin classifier η are inferred. Now, let us assume that we know the traffic phase of each test video clip. We employ the additional information of traffic phases for abnormality detection in test video clips by using the trained supervised topic models. Fig. 3 illustrates the flowchart of our proposed method for abnormality detection.

Using MedLDA with Dirichlet prior α , a large-margin classifier η , a set of topics β , for any given test video clip d and the corresponding label y_d , the topic mixture θ_d and topic z_d , we can write:

$$\begin{aligned} \min_{q, \alpha(\eta), \xi_d} & -E[\log p(\theta_d, z_d, w_d | \alpha, \beta)] \\ & - H(q(\theta_d, z_d)) + KL(q(\eta) || p_0(\eta)) + C \sum_{d=1}^D \xi_d \\ \text{s.t.} & E[\eta^T \mathbf{f}(y_d, \bar{z}_d) - \eta^T \mathbf{f}(y, \bar{z}_d)] \geq 1 - \xi_d \\ & \forall y \neq y_d, \xi_d \geq 0 \end{aligned} \quad (11)$$

Then, the likelihood is calculated as:

$$\begin{aligned} p(\theta_d, z_d, w_d, y_d | \alpha, \beta, \eta) & = p(\theta_d | \alpha) \times \\ & \left(\prod_{n=1}^N p(z_{dn} | \theta_d) p(w_{dn} | z_{dn}, \beta) \right) \times \\ & p(y_d | z_d, \eta) \end{aligned} \quad (12)$$

Obviously, a clip with a low likelihood is likely to include abnormalities. We will employ normalized log-likelihood in our experiments.

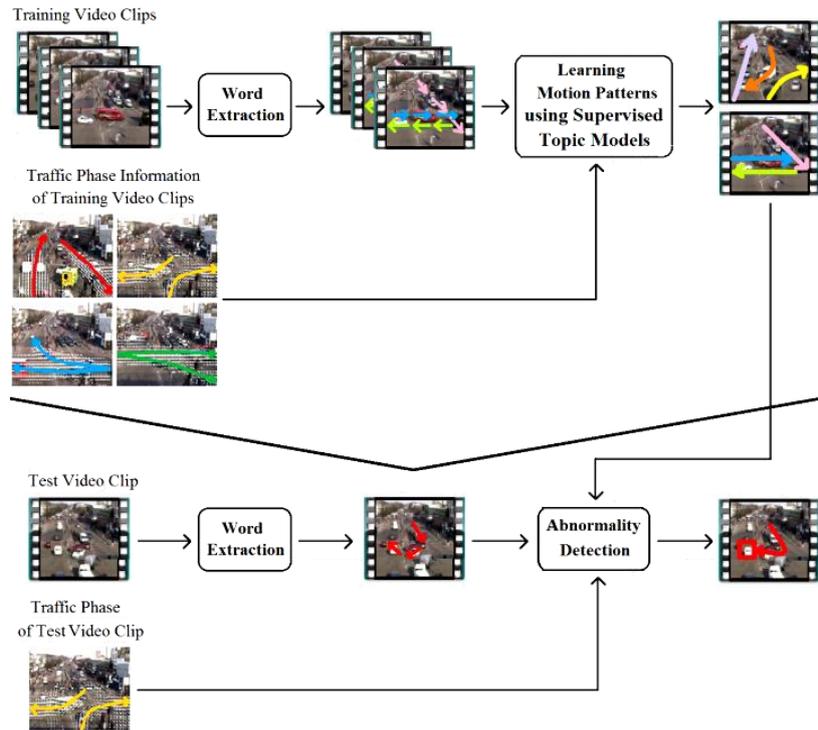


Fig. 3 Flowchart of our proposed method for abnormality detection.

Similarly, using MedSTC with a large-margin classifier η and a dictionary β , for a given test video clip d and its corresponding label y_d , the document code θ_d and word codes s_d , we can write:

$$\begin{aligned}
 \min_{\theta_d, s_d} & \sum_n (w_{dn} \log(s_{dn}^T \beta_n) - s_{dn}^T \beta_n) \\
 & + \sum_n (\gamma \|s_{dn} - \theta_d\|_2^2 + \rho \|s_{dn}\|_1) + \lambda \sum_d \|\theta_d\|_1 \\
 & + C \max_y [\Delta l(y_d, y) + \eta_y^T \theta_d - \eta_{y_d}^T \theta_d] \\
 \text{s.t.} & \quad \theta_d \geq 0; \quad s_{dn} \geq 0, n \in I_d
 \end{aligned} \quad (13)$$

Then, a sparse reconstruction cost (SRC) is calculated as:

$$\begin{aligned}
 f_{SRC} & = \sum_{n=1}^{|I_d|} (s_{d,n}^T \beta_n - w_{d,n} \ln(s_{d,n}^T \beta_n)) + \lambda \|\theta_d\|_1 \\
 & + \rho \sum_{n=1}^{|I_d|} \|s_{d,n}\|_1 + \gamma \sum_{n=1}^{|I_d|} \|s_{d,n} - \theta_d\|_2^2 \\
 & + \max_y [\Delta l(y_d, y) + F(y, \theta_d) - F(y_d, \theta_d)]
 \end{aligned} \quad (14)$$

Obviously, a clip with a high f_{SRC} tends to be with abnormalities.

4 Experimental Results

The output of the methods presented in this paper is tested using QMUL Junction dataset[‡]. This video has

[‡]http://www.eecs.qmul.ac.uk/~sgg/QMUL_Junction_Datasets/Junction/Junction.html

been recorded at 25 fps with a camera installed at a crossroad and its display resolution is 360×288 pixels. The Junction dataset includes 73 training clips and 39 test clips, where each clip has 12s length. Without any prior information, the number of topics is usually assigned with a large value so, here, the number of topics is given as $K=20$.

Stop lights control the QMUL Junction with 4 traffic phases, i.e. moving in upright lane, turn right and left with upright lane, and keep to the right or to the left. Lets suppose that we know the traffic phases of all training and test video clips so as to employ the method given in this paper to derive the motion patterns.

Figs. 4 and 5 depict sample motion patterns discovered by MedLDA and MedSTC, respectively, employing the traffic phase information. In addition, without using any supplementary information, motion patterns learned by LDA and STC are also added to these figures to make a comparison. As indicated, there are more plain semantic relations in the motion patterns learned by employing the traffic phase information.

Our proposed method is also tested on abnormal event detection. Eight abnormal events are found in 39 test clips captured by Junction dataset. These events are hazardous driving, not complying with traffic laws, causing interruption to traffic flow, uncommon moves like U-turns.

Receiver Operating Characteristic (ROC) curves for MedLDA and MedSTC are shown in Figs. 6 and 7, which are based on traffic phase information. Also, ROC curves for LDA and STC without employing any supplementary information are drawn in these figures to



Fig. 4 Motion patterns discovered by LDA and MedLDA.

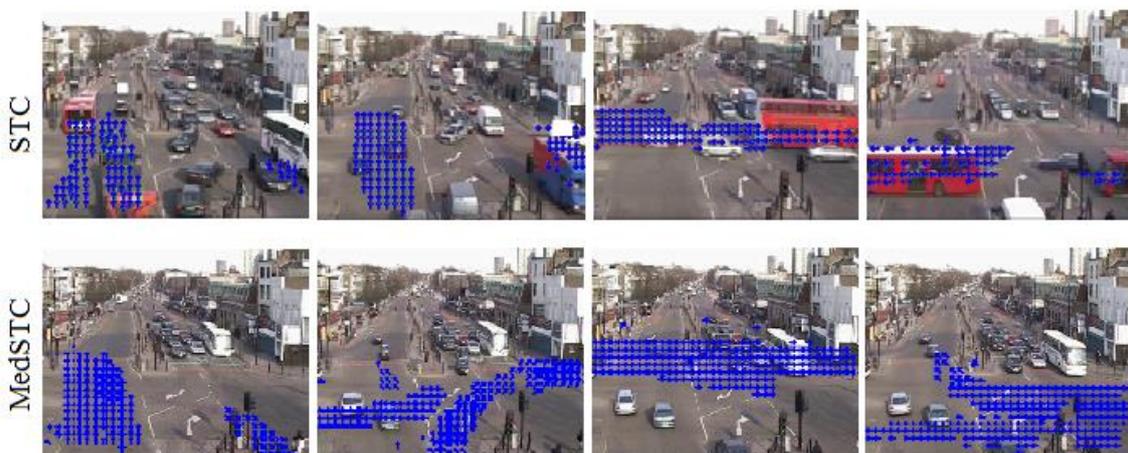


Fig. 5 Motion patterns discovered by STC and MedSTC.

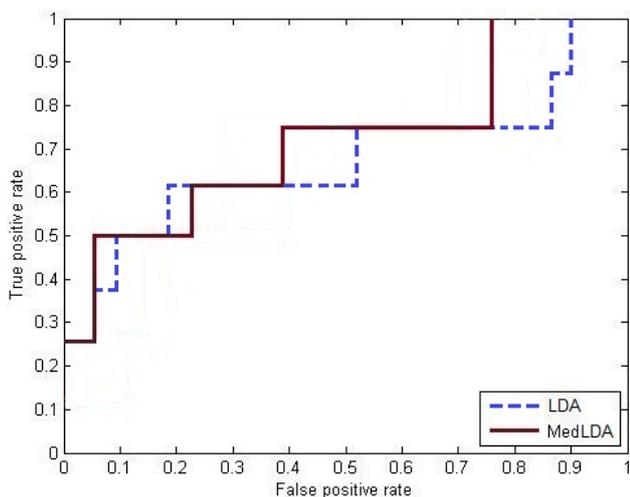


Fig. 6 ROC curves for abnormality detection using LDA and MedLDA.

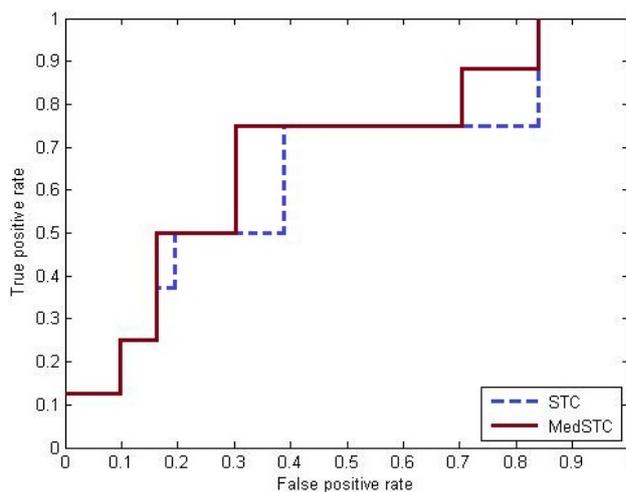


Fig. 7 ROC curves for abnormality detection using STC and MedSTC.

make a comparison. Based on the ROC curves shown in Figs. 6 and 7, the area under the ROC curves of LDA and STC are 71% and 69%, respectively. However, by using traffic phase information, the area under the ROC curves for MedLDA and MedSTC reach to 75% and

73%, respectively. This indicates that providing additional traffic phase information in the training phase results in more accurate motion pattern modeling, and hence, better performance in abnormality detection.

5 Computational Complexity

We analyzed the computational complexity of different methods, in learning motion patterns and abnormal event detection. Our experiments were performed on an Intel Core™ i7-4790 3.6 GHz CPU, 32 GB RAM, with Linux (Ubuntu 14.04) OS and C++ implementation. Using STC and MedSTC topic models, it takes about one second for training model on 73 12-second 360×288-pixel clips in QMUL Junction dataset. Testing 39 12-second clips with the same resolution and the same dataset, for abnormal event detection, requires about 2.5 second. For LDA and MedLDA topic models, training 73 video clips and testing 39 video clips take about 6.5 and 0.05 second, respectively. In comparison to STC, LDA has slower learning due to slow inference algorithm and solving various optimization problems.

6 Conclusion

In this paper, we address the problem of typical motion pattern discovery in traffic videos and its usage for abnormality detection. For this purpose, we assume that additional traffic information are assigned to the video clips. Based on such auxiliary information, we have formulated the problem of traffic motion pattern extraction and abnormality detection under supervised topic modeling framework. Experimental results show that incorporating side information makes supervised topic models, such as MedLDA and MedSTC, to discover more interpretable topic. Moreover, it leads to higher performance in abnormality detection. Compared to classic unsupervised topic models with no side information our method achieved more than 4% higher AUROC in abnormality detection on a widely-used traffic dataset.

References

- [1] W. Fu, J. Wang, H. Lu, and S. Ma, "Dynamic scene understanding by improved sparse topical coding," *Pattern Recognition*, Vol. 46, No. 7, pp. 1841–1850, 2013.
- [2] Y. Cai, H. Wang, X. Chen, and H. Jiang, "Trajectory-based anomalous behaviour detection for intelligent traffic surveillance," *IET Intelligent Transport Systems*, Vol. 9, No. 8, pp.810–816, 2015.
- [3] J. Ren, Y. Chen, L. Xin, J. Shi, B. Li, and Y. Liu, "Detecting and positioning of traffic incidents via video-based analysis of traffic states in a road segment," *IET Intelligent Transport Systems*, Vol. 10, No. 6, pp.428–437, 2016.
- [4] B. T. Morris and M. M. Trivedi, "Understanding vehicular traffic behavior from video: A survey of unsupervised approaches," *Journal of Electronic Imaging*, 22, No. 4, p. 041113, 2013.
- [5] X. Tang, X. L. Huang, S. Y. Sun, H. Dong, X. Zhang, Y. Gao, and N. Liu, "Intelligent recognition of traffic video based on mixture LDA model," in *International Conference on Machine Learning and Intelligent Communications*, Springer, Cham, pp. 356363, 2016.
- [6] V. Kaltsa, A. Briassouli, I. Kompatsiaris, and M. G. Strintzis, "Multiple hierarchical Dirichlet processes for anomaly detection in traffic," *Computer Vision and Image Understanding*, Vol. 169, pp.28–39, 2018.
- [7] L. Song, F. Jiang, Z. Shi, and A. Katsaggelos, "Understanding dynamic scenes by hierarchical motion pattern mining," in *IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, 2011.
- [8] J. Li, S. Gong, and T. Xiang, "Learning behavioural context," *International Journal Of Computer Vision*, Vol. 97, No. 3, pp. 276–304, 2012.
- [9] R. Emonet, J. Varadarajan, and J. M. Odobez, "Temporal analysis of motif mixtures using Dirichlet processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 1, pp. 140–156, 2014.
- [10] S. Rana, D. Phung, S. Pham, and S. Venkatesh, "Large-scale statistical modeling of motion patterns: a bayesian nonparametric approach," in *Indian Conference on Computer Vision, Graphics and Image Processing*, ACM, 2012.
- [11] T. Hofmann, "Probabilistic latent semantic analysis," in *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 289–296, 1999.
- [12] M. Blei, A. Y. Ng, M. I. Jordan, and J. Lafferty, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, Vol. 3, No. Jan, pp. 993–1022, 2003.
- [13] J. Zhu and E. Xing, "Sparse topical coding," in *Proceedings of the Twenty-Seventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 831–838, 2011.
- [14] Zhu J, Ahmed A, and Xing EP, "MedLDA: maximum margin supervised topic models for regression and classification," in *Proceedings of the 26th Annual International Conference on Machine Learning*, ACM, pp. 1257–1264, 2009.
- [15] D. Li, S. Somasundaran, and A. Chakraborty, "A combination of topic models with max-margin learning for relation detection," in *Proceedings of TextGraphs-6: Graph-based Methods for Natural Language Processing*, Association for Computational Linguistics, pp. 1–9, 2011.



P. Ahmadi received the degree of B.Sc. and M.Sc. both in Electrical Engineering from Isfahan University of Technology and Ph.D. from Sharif University of Technology, Tehran, Iran. Her research interests are mainly in the areas of signal and image processing, computer vision and pattern recognition. She is now an Assistant Professor at IT Research

Faculty, Iran Telecommunication Research Center.



I. Gholampour received his B.Sc., M.Sc. and Ph.D. in Electrical Engineering, all from Sharif University of Technology Tehran, Iran. His current interests include signal processing systems, machine vision, image and video analysis for intelligent traffic systems. He is now an Assistant Professor at Electronics Research Institute, Sharif University of

Technology.



© 2019 by the authors. Licensee IUST, Tehran, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license (<https://creativecommons.org/licenses/by-nc/4.0/>).