Conference or Workshop Item:

Isupova, O., Kuzin, D. orcid.org/0000-0003-3582-722X and Mihaylova, L. (2016) Anomaly detection in video with Bayesian nonparametrics. In: ICML2016 Anomaly Detection Workshop, June 24th, 2016, New York, USA.

Anomaly Detection in Video with Bayesian Nonparametrics

Olga Isupova O.ISUPOVA@SHEFFIELD.AC.UK
Danil Kuzin DKUZIN1@SHEFFIELD.AC.UK
Lyudmila Mihaylova L.S.MIHAYLOVA@SHEFFIELD.AC.UK
The University of Sheffield,
Western Bank, Sheffield, S10 2TN, UK

Abstract

A novel dynamic Bayesian nonparametric topic model for anomaly detection in video is proposed in this paper. Batch and online Gibbs samplers are developed for inference. The paper introduces a new abnormality measure for decision making. The proposed method is evaluated on both synthetic and real data. The comparison with a non-dynamic model shows the superiority of the proposed dynamic one in terms of the classification performance for anomaly detection.

1. Introduction

Topic modeling (Hofmann, 1999; Blei et al., 2003) is a promising approach for anomaly detection in video (Jeong et al., 2014; Varadarajan & Odobez, 2009; Mehran et al., 2009). This is an unsupervised method which means that there is no need to predict all kinds of abnormalities in advance and collect a labelled dataset for it. Topic modeling also provides additional information about typical motions and behaviours rather than just warns about abnormal events.

In the text mining application a topic model represents unlabelled documents as mixtures of topics where unknown topics are distributions over observed words. In conventional topic modeling documents are assumed to be independent. Although in some cases this assumption is not valid and different dynamic models are proposed in the literature (Blei & Lafferty, 2006; Ahmed & Xing, 2010; Hospedales et al., 2012; Kuettel et al., 2010; Pruteanu-Malinici et al., 2010; Srebro & Roweis, 2005; Zhang et al., 2010).

Presented at ICML2016 Anomaly Detection Workshop, New York, NY, USA, 2016.

In the video processing application short video clips are often treated as documents, local motion patterns are represented by topics. All motions in the real life last for some time hence topic mixtures in the successive documents are expected to be similar if the clips are sufficiently short.

We propose a dynamic nonparametric topic model for anomaly detection. Successive documents are encouraged to have similar topic mixtures.

Anomaly detection is an urgent task; the decision should be made as soon as possible. Batch and online Gibbs sampler is proposed in this paper. The online inference algorithm allows to estimate parameters for the current document with no need to rerun it on the previous ones. An abnormality measure for decision making is also proposed in the paper.

The paper is organised as follows. Section 2 defines visual words and documents while section 3 describes the proposed model. The inference and the whole framework are introduced in sections 4 and 5 respectively. Evaluation of the method using synthetic and real data is presented in section 6 followed by the conclusions in section 7.

2. Visual Features

The definitions of visual words and visual documents are essential for topic modeling application to video processing. A quantised direction (Figure 1) of an average optical flow vector (Horn & Schunck, 1981) over $N \times N$ pixels and its location form a visual word. Non-overlapping clips of the whole video sequence are treated as visual documents.

3. Proposed Model

Let $\mathbf{x}_{1:J} = \{\mathbf{x}_j\}_{j=1:J}$ denote a sequence of documents. A document $\mathbf{x}_j$ consists of N_j words x_{ji}: $\mathbf{x}_j = \{x_{ji}\}_{i=1:N_j}$. Documents are assumed to be mixtures of *topics* $\{\phi_k\}_{k=1:\infty}$, which are the latent distribution over

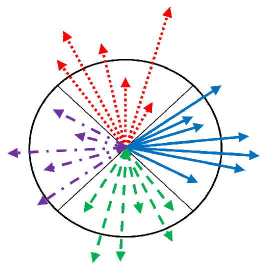

Figure 1. Optical flow direction quantisation. The four main directions are extracted — up, right, down and left, highlighted by the colour on the figure.

words. The number of possible topics is expected to go to infinity for an infinite amount of data.

3.1. Hierarchical Dirichlet Process Topic Model

A hierarchical Dirichlet process (HDP) (Teh et al., 2006) can represent mixture models with a potentially infinite number of mixture components.

A HDP can be represented in different ways, Chinese restaurant franchise (CRF) is reviewed here. Documents are considered as "restaurants" and words are considered as "customers" in this metaphor. The words in the documents form groups around "tables" and eat the same "dish", which corresponds to a topic, on one table. The set of the topics is shared among all the documents, that creates a "franchise" of the restaurants.

Denote a table assignment of the token i in the document j by t_{ji}, a topic assignment of the table t in the document j by k_{jt}. The following counts are used: n_{jt} for the number of words assigned to the table t in the document j and m_{jk} for the number of tables serving the topic k in the document j. Let dots in subscripts denote marginalisation over the corresponding dimension.

The generative process is the following. Each token i in the document j chooses one of the occupied tables with a probability proportional to a number n_{jt} of words already assigned to a table, or the token starts a new table with a probability proportional to a parameter α:

$$p(t_{ji} = t | t_{j1}, \dots, t_{ji-1}, \alpha) = \begin{cases} \dfrac{n_{jt}}{i-1+\alpha}, & \text{if } t = 1 : m_{j\cdot}; \\ \dfrac{\alpha}{i-1+\alpha}, & \text{if } t = t^{\text{new}}. \end{cases} \tag{1}$$

If a new table is started the topic should be assigned to it. It can be one of the used topics with a probability proportional to a number $m_{\cdot k}$ of tables having this topic among all the documents, or it can be a new topic with a probability proportional to a parameter γ:

$$p(k_{jt^{\text{new}}} = k | k_{11}, \dots, k_{jt-1}, \gamma) = \begin{cases} \dfrac{m_{\cdot k}}{m_{\cdot\cdot} + \gamma}, & \text{if } k = 1 : K; \\ \dfrac{\gamma}{m_{\cdot\cdot} + \gamma}, & \text{if } k = k^{\text{new}}, \end{cases} \tag{2}$$

where K is a number of topics used so far. In the case of a new topic it is sampled from the base measure H.

The word x_{ji} for the topic i in the document j assigned to the table t_{ji} is sampled from the topic $k_{jt_{ji}}$ served on this table:

$$x_{jt} \sim \text{Mult}(\phi_{k_{jt_{ji}}}). \tag{3}$$

3.2. Dynamic Hierarchical Dirichlet Process Topic Model

Exchangeability of documents and words is an essential assumption in the HDP, which means that the joint probability of the data is independent of the order of the documents and words. Although in video processing this assumption is not reasonable. Motions last for some time and it is expected that the topic mixture of the current document is similar to the topic mixture in the previous one. However, the words inside documents are still exchangeable.

The dynamic extension of the HDP topic model is proposed in this paper to take into account this intuition. The probability of the topic k being assigned to one of the tables in the document j explicitly depends on the usage of this topic in the current and previous documents $m_{jk} + m_{j-1k}$. The topic distribution of the current document is hence encouraged to be similar to the topic distribution of the previous one.

The proposed model assumes the following generative process. A table assignment for a token remains unchanged (1). A topic for a new table in the document j is assigned to one of the used topics k with a probability proportional to the sum of the number of tables serving this topic in the current and previous documents $m_{jk} + m_{j-1k}$ and the weighted number of tables among all the documents that have this topic $\delta\, m_{\cdot k}$, where δ is a parameter of the model, or it is assigned to a new topic with a probability proportional to the parameter γ:

$$p(k_{jt} = k | k_{11}, \dots, k_{jt-1}, \gamma) = \begin{cases} \dfrac{m_{jk} + m_{j-1k} + \delta m_{\cdot k}}{m_{j\cdot} + m_{j-1\cdot} + \delta m_{\cdot\cdot} + \gamma}, & \text{if } k = 1 : K; \\ \dfrac{\gamma}{m_{j\cdot} + m_{j-1\cdot} + \delta m_{\cdot\cdot} + \gamma}, & \text{it } k = k^{\text{new}}. \end{cases} \tag{4}$$

The word x_{ji} is sampled as in the HDP from the corresponding topic as defined in (3).

Table 1. AUC results.

DATA SET	DYNAMIC HDP	HDP	"TRUE" MODEL
SYNTHETIC	0.7118	0.4751	0.7280
QMUL	0.7100	0.4644	—

4. Inference

Conventional inference algorithms are batch algorithms, i.e. they process the whole dataset, which is computationally intractable for large or stream datasets. Online algorithms work sequentially, one data point at a time. We propose a combination of offline and online inference for our model.

We use Gibbs sampling (Geman & Geman, 1984). The hidden variables $\mathbf{t} = \{t_{ji}\}_{j=1:J,i=1:N_j}$ and $\mathbf{k} = \{k_{jt}\}_{j=1:J,t=1:m_{j\cdot}}$ are sampled from their conditional distributions.

The batch Gibbs sampler is run for the training set of the documents. After this training stage the global estimates of the topics ϕ_k and the counts $m_{\cdot k}$ for all k are stored and used for the online inference of the testing documents. For each testing document the online Gibbs sampler is run to sample table assignments and topic assignments for this document only. The online Gibbs sampler updates the local counts m_{jk}. After the Gibbs sampler converges, the global variables ϕ_k and $m_{\cdot k}$ are updated with the information obtained by the new document.

5. Anomaly Detection

Anomaly detection can be done within the probabilistic framework using topic modeling. In this framework the data point is assumed to be abnormal if it has a low value of likelihood, i.e. the learnt model cannot explain the current observation because something atypical is happening. We use the predictive likelihood estimated as a harmonic mean (Griffiths & Steyvers, 2004) and normalised by the length N_j of the document for anomaly detection.

6. Experiments

The proposed method[1] is applied for anomaly detection on synthetic and real data. We compare it with the method based on the HDP topic model (for the batch Gibbs sampler of the HDP topic model the implementation by Chong Wang is used[2]). For the quantitative comparison the area (AUC) under the receiver operating characteristic (ROC) curve (Murphy, 2012) for abnormality classification accuracy is used.

[1]https://github.com/OlgaIsupova/dynamic-hdp
[2]https://github.com/Blei-Lab/hdp

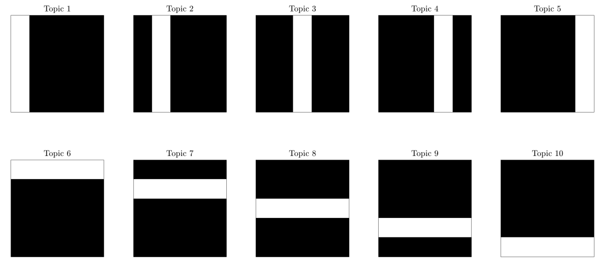

Figure 2. Graphical representation of the topics in the synthetic dataset. There are 25 words, organised into a 5×5 matrix, where a word corresponds to a cell in this matrix. The topics are represented as the coloured matrices, where the colour of the cell indicates the probability of the corresponding word in a given topic, the lighter the colour the higher the probability value is.

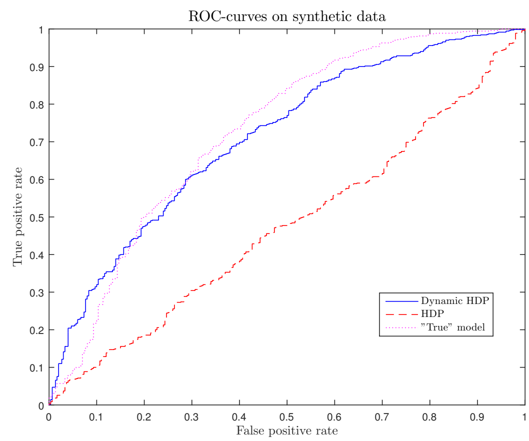

Figure 3. The ROC-curves for the synthetic data.

6.1. Synthetic Data

The "bar" data introduced in (Griffiths & Steyvers, 2004) is used. The vocabulary consists of $V = 25$ words, organised into a 5×5 matrix. The topics $\{\phi_k\}_{k=1}^{10}$ form vertical and horizontal bars in the matrix (Figure 2).

Within the testing dataset we generate some "abnormal" documents where topics are chosen uniformly from the set of all the topics except those used in the previous documents. The data generated in such a way contradicts the main model assumption that the topic mixtures of the successive documents should be similar.

Figure 3 presents the obtained ROC-curves for anomaly detection. For the reference we also show the ROC-curve for the "true" model, i.e. the model with the true topics ϕ_k and

(a) (b) (c) (d)

Figure 4. QMUL-junction dataset snapshots. (a) is an example of a normal motion, (b) is an example of jay-walking abnormality, (c) is an example of a car moving on the wrong lane in the opposite to normal direction, (d) is an example an emergency service car disrupting a normal traffic flow.

the true table and topic assignments $\mathbf{t}$ and $\mathbf{k}$. This model represents the one that can perfectly restore all the latent variables. The corresponding AUC values are in Table 1. The proposed dynamic HDP shows the anomaly detection performance competitive to the "true" model.

6.2. Real Data

We also test the algorithms on the QMUL-junction real video data (Hospedales et al., 2012) captured a busy road junction (Figure 4(a)). 5 out of 45 minutes of the video sequence is used as a training dataset for offline Gibbs sampler.

For the ground truth reference the data is labelled as normal and abnormal, where abnormal event examples are jay-walking (Figure 4(b)), driving wrong direction (Figure 4(c)), disruption in traffic flow (Figure 4(d)).

Figure 5 presents the ROC-curves while Table 1 contains the corresponding AUC values. The experiment confirms that the dynamics consideration in a topic model improves the anomaly detection performance.

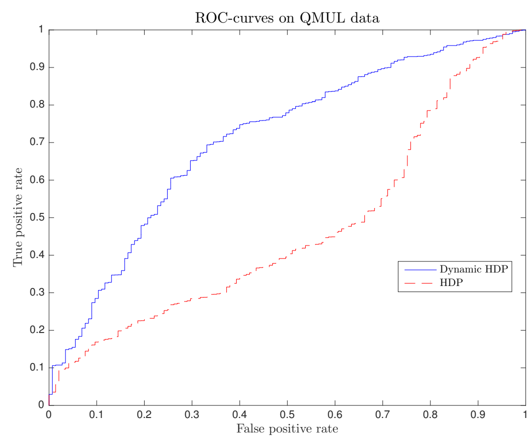

Figure 5. The ROC-curves for the QMUL data.

7. Conclusions

A novel Bayesian nonparametric dynamic topic model is proposed in this paper. Batch and online inference algorithms are designed. Anomaly detection in video is considered as an application of the model for which we propose an abnormality measure. The experimental results both on synthetic and real data show that the proposed dynamic topic model improves the anomaly detection performance in comparison to the non-dynamic model.

Acknowledgments

This work is accepted to the 19th International Conference on Information Fusion 2016. Olga Isupova and Lyudmila Mihaylova acknowledge the support from the EC Seventh Framework Programme [FP7 2013-2017] TRAcking in compleX sensor systems (TRAX) Grant agreement no.: 607400. Lyudmila Mihaylova acknowledges also the support from the UK Engineering and Physical Sciences Research Council (EPSRC) via the Bayesian Tracking and Reasoning over Time (BTaRoT) grant EP/K021516/1.

References

Ahmed, Amr and Xing, Eric. Timeline: A dynamic hierarchical Dirichlet process model for recovering birth/death and evolution of topics in text stream. In *Proceedings of the Twenty-Sixth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-10)*, pp. 20–29, Corvallis, Oregon, 2010. AUAI Press.

Blei, David M. and Lafferty, John D. Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, pp. 113–120, New York, NY, USA, 2006. ACM.

Blei, David M., Ng, Andrew Y., and Jordan, Michael I. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, March 2003.

Geman, Stuart and Geman, Donald. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):721–741, 1984.

Griffiths, Thomas L. and Steyvers, Mark. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(1):5228–5235, 2004.

Hofmann, Thomas. Probabilistic latent semantic indexing. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '99, pp. 50–57, New York, NY, USA, 1999. ACM.

Horn, Berthold K and Schunck, Brian G. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

Hospedales, Timothy, Gong, Shaogang, and Xiang, Tao. Video behaviour mining using a dynamic topic model. *International Journal of Computer Vision*, 98(3):303–323, 2012.

Jeong, Hawook, Yoo, Youngjoon, Yi, Kwang Moo, and Choi, Jin Young. Two-stage online inference model for traffic pattern analysis and anomaly detection. *Machine Vision and Applications*, 25(6):1501–1517, 2014.

Kuettel, D., Breitenstein, M.D., Van Gool, L., and Ferrari, V. What's going on? Discovering spatio-temporal dependencies in dynamic scenes. In *Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1951–1958, June 2010.

Mehran, Ramin, Oyama, Alexis, and Shah, Mubarak. Abnormal crowd behavior detection using social force model. In *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 935–942. IEEE, June 2009.

Murphy, Kevin P. *Machine learning: a probabilistic perspective*. MIT press, 2012.

Pruteanu-Malinici, I., Ren, Lu, Paisley, J., Wang, E., and Carin, L. Hierarchical Bayesian modeling of topics in time-stamped documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):996–1011, June 2010.

Srebro, N. and Roweis, S. Time-varying topic models using dependent Dirichlet processes. Technical report, Technical Report, Departament of computer science, University of Toronto, 2005.

Teh, Yee Whye, Jordan, Michael I, Beal, Matthew J, and Blei, David M. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.

Varadarajan, J. and Odobez, J. Topic models for scene analysis and abnormality detection. In *Proceedings of the 2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 1338–1345, Sept 2009.

Zhang, Jianwen, Song, Yangqiu, Zhang, Changshui, and Liu, Shixia. Evolutionary hierarchical Dirichlet processes for multiple correlated time-varying corpora. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1079–1088. ACM, 2010.