Abnormal behaviour detection in video using topic modeling

Olga Isupova[1], Danil Kuzin[1], Lyudmila Mihaylova[1]

[1](Department of Automatic Control and Systems Engineering, Faculty of Engineering, The University of Sheffield)

Abstract

The growth of the number of surveillance systems makes it is impossible to process data by human operators thereby autonomous algorithms are required in a decision-making procedure. A novel dynamic topic modeling approach for abnormal behaviour detection in video is proposed. Activities and behaviours in the scene are described by the topic model where temporal dynamics for behaviours is assumed. Here we implement Expectation-Maximisation algorithm for inference in the model and show in the experiments that it outperforms the Gibbs sampling inference scheme that is originally proposed in [1].

Keywords Abnormal behaviour; computer vision; Expectation-Maximisation algorithm; Gibbs sampling; topic modeling

1. INTRODUCTION

The abnormal behaviour detection research area has become very attractive over the last decade. The formal definition of abnormality is one of the challenges. Abnormal behaviour can be determined as that not similar to the normal behaviour patterns extracted during a training stage [2-4]. The other approach is to consider a statistically rare event as abnormal. The statistical models are built, including one-class classifiers such as one-class Support Vector Machine [5] or topic models [6].

Topic modeling was originally proposed for text mining [7, 8]. The documents are represented as distributions over non-observed topics where the topics are distributions over words.

The Markov Clustering Topic Model (MCTM) is proposed in [1]. An abstract behaviour class is assumed determining an activity distribution for each video clip. We propose a novel inference scheme for similar generative model based on the maximum likelihood approach via the Expectation-Maximisation (EM) algorithm. The decision about abnormality is made calculating likelihood of test clips.

The rest is organised as follows. Visual word representation description is provided in Section 2. The review of the MCTM [1] is presented in Section 3. Section 4 describes the proposed framework. There are experimental results in Section 5. The conclusion is presented in Section 6.

2. VISUAL FEATURES

Visual features are calculated based on optical flow. The quantisation over a spatial domain is made by dividing frames into small cells. The motion is quantised into four directions. A cell position and its direction form a visual word. The short video clips are treated as documents.

3. MARKOV CLUSTERING TOPIC MODEL

Two novelties are proposed in the MCTM [1] compared to the original topic model [8]. Firstly, the topic distribution for each document (video clip) is chosen from the class of such distributions called *behaviours*. The visual word distributions – the topics, explain simple *actions* while behaviours explain complex interactions within the scene. Secondly, the temporal dynamics for behaviours is assumed modelled as a Markov chain.

Consider an example. There is a fixed camera on a road junction regulated by traffic lights. Behaviour in this case corresponds to a traffic light regime as these regimes have temporal dynamics and explain all the actions happening within the scene. An action here may be motion of the pedestrians on a crosswalk.

Word distributions within the topics, topic distributions within behaviours and behaviour dynamics are modelled as multinomial distributions. For all multinomial distributions Dirichlet prior is assumed. Inference is made by the collapsed Gibbs sampling scheme.

4. THE PROPOSED APPROACH

We propose a new inference scheme for the model denoted as EM-MCTM.

The generative model for the EM-MCTM is the same as for MCTM except that the algorithm does not rely on prior for multinomial distributions (the graphical model can be found in Figure 1).

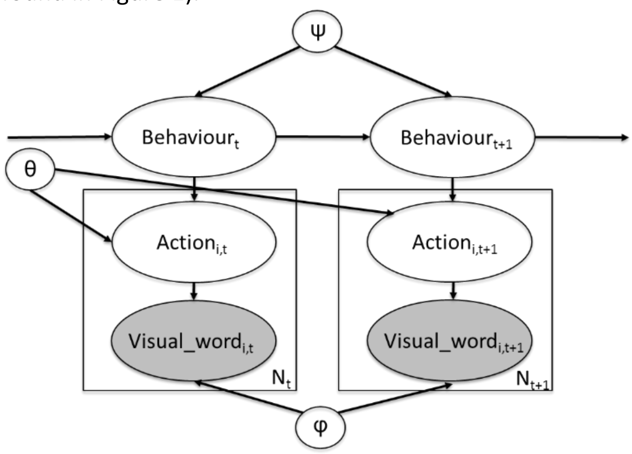

Figure 1. The graphical model of the proposed model

The model parameters are estimated with the maximum likelihood approach. The EM-algorithm (9) is applied to the optimisation problem.

A new observation likelihood is calculated using the estimates of the model parameters. If the likelihood is less than a threshold the observation is labelled as abnormal otherwise as normal.

5. EXPERIMENTS

The proposed EM-MCTM approach is applied for abnormality detection and compared with the MCTM proposed in [8] (denoted later as GS-MCTM). The experiments are made with both synthetic and real data.

We generate numerical synthetic data with some fixed model parameters. We also generate some `abnormal' data which cannot be explained with the given model parameters. The classification results for the synthetic data can be found in Table 1.

Table 1. The classification results for synthetic data

Quality measure	GS-MCTM	EM-MCTM
Error percentage	0.1503	0.1387
F-measure (10)	0.8961	0.9035

We used the University of Minnesota (UMN) video dataset for detection of unusual crowd activity [11]. The dataset is consists of 3 scenes and has a ground truth for the abnormality classification. The results of the classification for the UMN dataset can be found in Table 2.

Table 2. The classification results for the UMN dataset

Scene	Quality measure	GS-MCTM	EM-MCTM
First	**Error percentage**	0.0705	0.0385
	F-measure (10)	0.9573	0.9763
Second	**Error percentage**	0.2599	0.2528
	F-measure (10)	0.8402	0.8485
Third	**Error percentage**	0.0776	0.0750
	F-measure (10)	0.9574	0.9591

One can notice that our proposed method outperforms the one, based on the Gibbs sampling scheme, both with the synthetic and real data.

6. CONCLUSION

A novel dynamic topic model inference scheme is designed. A topic model explains typical actions and behaviours of a scene in video. Abnormal behaviour detection can be performed within the developed framework. A video clip having a small likelihood according to the built topic model is treated as abnormal. The maximum likelihood approach is implemented for inference. The maximisation is performed using the EM-algorithm while the Gibbs sampling scheme is originally proposed in [1]. The experiments both on synthetic and real data show that the proposed algorithm outperforms the one using Gibbs sampling inference.

7. ACKNOWLEDGEMENTS

The authors would like to thank the support from the EC Seventh Framework Programme [FP7 2013-2017] TRAcking in compleX sensor systems (TRAX) Grant agreement no.: 607400.

REFERENCES

1. Hospedales T, Gong S, Xiang T. Video behaviour mining using a dynamic topic model. International Journal of Computer Vision. 2012, 98(3): p. 303-323
2. Yen S-H, Wang C-H. Abnormal event detection using HOSF. In: 2013 International Conference on IT Convergence and Security (IC-ITCS); 2013 Dec 16-18; Macao. IEEE, 2013. p. 1-4
3. Ourivach K, Gharti S, Dailey M N. Incremental behaviour modeling and suspicious activity detection. Pattern Recognition. 2013, 46(3): p. 671-680
4. Cong Y, Yuan J, Liu J. Abnormal event detection in crowded scenes using sparse representation. Pattern Recognition. 2013, 46(7): p.1851-1864
5. Cheng K-W, Chen Y-T, Fang W-H. Abnormal crowd behavior detection and localization using maximum sub-sequence search. In: Proceedings of the 4th ACM/IEEE International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Stream; 2013; Barcelona, Spain. New York, NY, USA: ACM, 2013. p. 49-58
6. Mehran R, Oyama A, Shah M. Abnormal crowd behaviour detection using social force model. In: IEEE Conference on Computer Vision and Pattern Recognition; 2009 Jun 20-25; Miami, FL. IEEE, 2009. p. 935-942
7. Hofmann T. Probabilistic latent semantic indexing. In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrival, ser. SIGIR'99; 1999; Berkeley, USA. New York, NY, USA: ACM, 1999. p. 50-57
8. Blei D, Ng A, Jordan M. Latent Dirichlet Allocation. Journal of Machine Learning Research. 2003, 3: p. 993-1022
9. Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society. Series B (Methodological). 1977, 39(1): p. 1-38
10. Manning C, Raghavan P, Schutze H. Introduction to Information Retrieval. New York, NY, USA: Cambridge University Press; 2008
11. University of Minnesota dataset for detection of unusual crowd activity [Online]. Available from: http://mha.cs.umn.edu/proj_events.shtml#crowd