Learning Rare Behaviours

Jian Li, Timothy M Hospedales, Shaogang Gong, Tao Xiang

School of EECS, Queen Mary University of London
Mile End Road, London, E1 4NS, UK
Email: {jianli, tmh, sgg, txiang}@dcs.qmul.ac.uk

Abstract. We present a novel approach to detect and classify rare behaviours which are visually subtle and occur sparsely in the presence of overwhelming typical behaviours. We treat this as a weakly supervised classification problem and propose a novel topic model: Multi-Class Delta Latent Dirichlet Allocation which learns to model rare behaviours from a few weakly labelled videos as well as typical behaviours from uninteresting videos by collaboratively sharing features among all classes of footage. The learned model is able to accurately classify unseen data. We further explore a novel method for detecting unknown rare behaviours in unseen data by synthesising new plausible topics to hypothesise any potential behavioural conflicts. Extensive validation using both simulated and real-world CCTV video data demonstrates the superior performance of the proposed framework compared to conventional unsupervised detection and supervised classification approaches.

1 Introduction

Detecting rare behaviours in videos of dynamic scenes is important for video behaviour analysis. Existing studies usually treat this problem as either an outlier detection problem, flagging any behaviour which is badly explained by a model representing typical behaviours [1–5]; or a supervised classification problem when examples of rare behaviours are known and available for training an explicit model of rare versus typical behaviours [6, 7]. For both, the real challenge in model building is that visual evidence for rare behaviours is subtle and under-represented, which makes them neither distinctive enough to be detected among many concurrently occurring typical behaviours, nor repeated enough to be modelled precisely. Examples shown in Fig. 1 illustrate the difficulties of this problem. Two rare behaviours of single object (rare turns) are highlighted by red arrows, surrounded and overwhelmed by typical behaviours (yellow arrows) of dozens of other objects. In addition, both these rare turns are short in duration with few frames containing the crucial motion features that distinguish them from typical behaviours. Moreover, as rare behaviours, very few samples are available, making it difficult to model them reliably. All these issues cause both existing supervised and unsupervised detection models to fail.

In this work, we propose a novel approach to solve the problem of rare behaviour modelling and detection by treating it as a weakly supervised classification problem. In particular, we wish to exploit *weak* labelling of video footage as

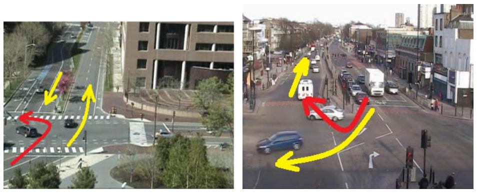


Fig. 1. Rare behaviours (red) may have weak visual evidence and usually co-occur with other typical behaviours (yellow).

additional information about any rare behaviours. For example, whilst the majority of daily CCTV recordings may be labelled as ‘uninteresting’ by default, a few clips may also be labelled as being interesting, containing rare occurrences of some interesting behaviours such as ‘U-turn’, ‘Jay-Walking’ and ‘tailgating’ but without exact information about where and what are the triggering rare behaviours. The challenge is how to quantify and utilise such weak labelling information, which is mostly qualitative (or binary), in order to overcome under-representation of those rare behaviours in model learning. To that end, given a set of such weakly labelled video footage, we introduce a novel Multi-Class Delta Latent Dirichlet Allocation (MC-ΔLDA) model, which uses weak supervision to jointly learn a model for both typical and rare behaviours with a partially shared representation. Online behaviour classification is performed by inference in a learned MC-ΔLDA model, which requires accurately computing marginal likelihoods. To this end, we formulate an importance sampler to estimate the marginal likelihood using a variational mean field approximation to the optimal proposal. Our framework exploits the benefits of both robust generative topic modelling and supervised classification. Importantly, this is without imposing the cost of and assuming consistency in fully supervised labelling. The proposed model permits us to learn to detect and classify different types of rare behaviours, even when they are visually subtle and buried in cluttering typical behaviours.

In contrast to supervised discriminative classifiers, our method retains the generative model potential to detect completely new and unseen rare behaviours as outliers due to their low likelihood – but with the already identified caveats of subtle and buried rare behaviours being challenging to detect as outliers. As a second contribution, we show how our framework can alternatively exploit very general prior knowledge about a given scene to *predict* unseen rare behaviours, and therefore detect them within the same MC-ΔLDA classification framework.

1.1 Related Work

Rare behaviour detection is typically formulated as outlier detection against a learned generative model of typical behaviours. For example, previous efforts have learned a dynamic Bayesian network (DBN) [1] or Gaussian mixture model (GMM) [2] from unlabelled video data, and threshold the likelihood of test data under these models to detect rare behaviour. These models however suffer from lack of robustness to noise and clutter in crowded scenes. Recently a number of studies on modelling complex scenes have instead exploited statistical topic models such as Latent Dirichlet Allocation (LDA) [8] to cluster low level features into topics – or activities in video – for robust modelling of behaviour [3–5]. Topic models aim to automatically discover meaningful activities via co-occurrence of visual words. They can deal with simultaneous interaction of multiple objects or activities. Despite these advantages, all unsupervised topic models [3–5] for rare behaviour detection must compute how well new examples can be explained by the learned typical behaviour model. This exposes their key limitation: unsupervised topic models are only sensitive to rare behaviours which are visually very distinct from the majority. That is, visually subtle rare behaviours (i.e. sharing substantial visual words of typical behaviours) are usually overwhelmed (explained away) by the more obvious and common behaviours, and therefore cannot be detected. Moreover, all outlier detection based approaches [1–5] have no mechanism to classify different types of rare behaviours.

Alternatively, supervised classification methods have also been applied to classify behaviours given sufficient and unbiased labelled training data [6, 7]. However, they are intrinsically limited in modelling rare behaviours due to under-represented rare data, making it difficult to build a good decision boundary. More importantly, even in a video clip with rare behaviours, the vast majority of observations are typical. This means that for a supervised classifier, the majority of video features in the clip are irrelevant without very specific, expensive and hard-wired supervision. In practice, weakly-supervised and in particular multi-instance learning (MIL) [9, 10] is preferred. Different from existing MIL methods which are based on discriminative models such as SVM [10] or boosting [9], our MC-ΔLDA is a generative model which retains the robustness of generative topic modelling whilst exploiting weak supervision (cheap and scalable) for both rare and typical behaviour classification in a single framework.

MC-ΔLDA intrinsically differs from Supervised Topic Models (sLDA) [11] in that it uses weak supervision to control the existence of a particular topic in a video clip instead of controlling how all topics are mixed, as in sLDA. More importantly, MC-ΔLDA is a classifier whereas sLDA was used for regression and generalising it to classification is non-trivial. Our algorithm is able to learn joint models of both typical and rare behaviours even if they co-exist. MC-ΔLDA is a generalisation and completion of the ΔLDA model proposed in [12]. ΔLDA was used for understanding code bugs in computer programs, but without an inference framework for the labels of unseen documents, ΔLDA cannot be used for classification. We extend ΔLDA (two-class only) by generalising it to MC-ΔLDA, which can jointly learn multi-class rare behaviours. Moreover, we complete the model by formulating an efficient and accurate inference algorithm to enable MC-ΔLDA to classify known rare behaviours in unseen data. We further explore MC-ΔLDA for detecting completely unknown rare behaviours through synthesising topics. Learning to classify by synthesis has been studied before for learning outlier detectors using classification models [13, 14]. However, such models synthesise without constraint, so predicted instances can be invalid and also risk double-counting evidence. Our MC-ΔLDA differs significantly in that it explores constrained synthesis under prior contextual knowledge in order to generate hypotheses that can best explain unknown but plausible rare behaviours.

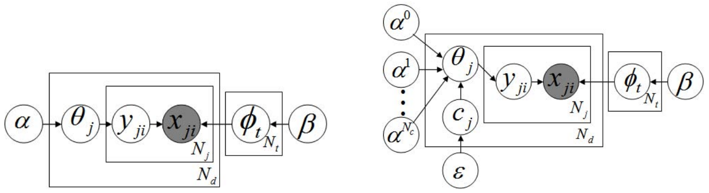
We extensively evaluate MC-ΔLDA using video data captured from two busy traffic scenes, and show its superiority for both classification and detection of known and unknown rare behaviours compared to existing supervised and unsupervised models.

2 Classify and Detect Rare Behaviours by MC-ΔLDA

2.1 Video Clip Representation

Training a topic model requires a bag of words representation of video. To create visual words, we first segment a scene into $N_a \times N_b$ equal size cells. Within each cell, we compute optical flow vectors for each pixel and the average vector for the cell. We quantise the cell motion into one of N_m directions. This results in a codebook of size $N_w = N_a \times N_b \times N_m$. To create visual documents, we temporally segment a video into N_d non-overlapping clips, and a document consists of the set of quantised words in that clip. In this work, a corpus consists of N_d documents $\mathbf{D} = \{\mathbf{x}_j\}_{j=1}^{N_d}$ each of which is a bag of N_j words $\mathbf{x}_j = \{x_{j,i}\}_{i=1}^{N_j}$.

2.2 Multi-Class Delta Latent Dirichlet Allocation (MC-ΔLDA)



(a) LDA (b) MC-ΔLDA

Fig. 2. LDA vs. MC-ΔLDA model structures.

Standard topic models such as LDA [8] (Fig. 2 (a)) model each instance (clip) as being derived from a bag of topics drawn from a fixed (and usually uniform) set of proportions. In MC-ΔLDA, we wish to constrain the topic proportions non-uniformly and on a per-clip basis. This will enable some topics to be shared (perpetually ongoing typical behaviours) and some to be uniquely associated with particular interesting classes, thereby representing the unique aspects of that behaviour. The new model structure is shown in Fig. 2 (b). Each clip $\mathbf{x}_j$ is assumed to be represented by a bag of words drawn from a mixture of $N_t = \sum_{c=0}^{N_c} N_{t,c}$ possible topics, where $N_{t,c}$ is the number of topics allocated to behaviour c. Let T_c be the $N_{t,c}$ element list of topics for behaviour c. Each clip is assumed to belong to a specific behaviour class $c_j \in 0, 1, \cdots, c, \cdots, N_c$. $c_j = 0$ indicates that clip $\mathbf{x}_j$ contains only typical behaviours; and $c_j = 1, \cdots, c, \cdots, N_c$ indicates that clip $\mathbf{x}_j$ contains both typical behaviours and also a type c rare behaviour (in unknown proportions and at unknown locations). This is enforced by defining a class specific prior over topics $\alpha^c = \{\alpha_t^c\}_{t=1}^{N_t}$ where elements $t \notin T_0 \bigcup T_c$ are constrained to be zero. In this study we set all non-zero elements of α^c to 0.5, and β is learned with Gibbs-expectation maximization [15]. The generative process of MC-ΔLDA is summarized as follows:

1. Draw a Dirichlet word-topic distribution $\phi_t \sim \mathrm{Dir}(\beta)$ for every topic t;
2. For each document j:
   (a) Draw a class label $c_j \sim \mathrm{Multi}(\varepsilon)$;
   (b) Given label c_j, draw a constrained topic distribution $\theta_j \sim \mathrm{Dir}(\alpha^{c_j})$;
   (c) Draw a topic $y_{j,i}$ for each word i from multinomial $y_{j,i} \sim \mathrm{Multi}(\theta_j)$;
   (d) Sample a word $x_{j,i}$ according to $x_{j,i} \sim \mathrm{Multi}(\phi_{y_{j,i}})$.

During the training phase, we assume weak supervision is provided in the form of labels c_j, and the goal is to determine the topics for each clip (i.e. θ_j and $\mathbf{y}_j$) and their visual representation (i.e. Φ), including ongoing typical behaviours, and unique rare behaviours. In the testing phase, the class label of an unseen document $\mathbf{x}^*$ is not available and our objective is to infer its class, $p(c|\mathbf{x}^*)$.

2.3 Learning: Modelling Rare Behaviours with Weak Supervision

The full joint probability of a document j in MC-ΔLDA is

$$p(\mathbf{x}_j, \mathbf{y}_j, \theta_j, \Phi, c|\alpha, \beta, \varepsilon) = \prod_i p(x_{j,i}|y_{j,i}, \Phi)p(y_{j,i}|\theta_j)p(\theta_j|\alpha, c)p(c|\varepsilon)p(\Phi|\beta). \quad (1)$$

As for standard LDA, exact learning in our model is intractable. However a collapsed Gibbs sampler can be derived to sample the topic posterior $p(\mathbf{y}|\mathbf{x}, c, \alpha, \beta)$ (now additionally conditioned on the current class c) leading to the update

$$p(y_{j,i}|\ \mathbf{y}_{j,-i}, \mathbf{x}, c, \alpha, \beta) \propto \frac{n_{x,y}^{-i} + \beta}{\sum_x n_{x,y} + \beta} \frac{n_{y,d}^{-i} + \alpha_y^{c_j}}{\sum_y n_{y,d}^{-i} + \alpha_y^{c_j}}. \quad (2)$$

Here $\mathbf{y}_{j,-i}$ indicates all topics except the token i; $n_{x,y}^{-i}$ indicates the counts of topic y being assigned to word x, excluding the current item i; and $n_{y,d}^{-i}$ indicates the count of topic y occurring in the current document d. These counts are also used to point estimate Dirichlet parameters θ and Φ by their mean, e.g.

$$\hat{\Phi}_{x,y} = \frac{n_{x,y} + \beta}{\sum_x n_{x,y} + \beta}. \quad (3)$$

Note that in contrast to standard LDA, the class constrained priors α^{c_j} in Eq. (2) enforce that the topic posterior $y_{j,i}$ (for all words i in clip j) is zero except for topics permitted by the current class c_j $(T_0 \bigcup T_{c_j})$. Topics $T_{c=0}$ will be well constrained by the abundant typical data. However, all clips of some interesting class $c > 0$ may use unique extra topics T_c in their representation. These will therefore come to represent the unique aspects of interesting class c.

2.4 Inference: Classification of Known Rare Behaviours

The learning problem formulated above exploited labelled behaviours c. For online classification of new/unseen test data $\mathbf{x}^*$, the goal is to estimate $c = \mathrm{argmax}_c p(c|\mathbf{x}^*, \mathbf{D}, \alpha, \beta, \varepsilon)$ where

$$p(c|\mathbf{x}^*, \mathbf{D}, \alpha, \beta, \varepsilon) \propto p(\mathbf{x}^*|\mathbf{D}, c, \alpha, \beta)p(c|\varepsilon), \quad (4)$$

$$p(\mathbf{x}^*|\mathbf{D}, c, \alpha, \beta, \varepsilon) = \int\int \sum_{\mathbf{y}} p(\mathbf{x}^*|\mathbf{y}, \Phi)p(\mathbf{y}|\theta)p(\Phi|\mathbf{D})p(\theta|\alpha, c)d\theta d\Phi, \quad (5)$$

The challenge is then to accurately estimate the marginal likelihood in Eq. (5). Firstly, we approximate $p(\Phi|D) = \delta_{\hat{\Phi}}(\Phi)$, taking Φ to be fixed at single point estimate $\hat{\Phi}$ from the final Gibbs sample during learning (Eq. (3)). We point out that reliably computing the marginal likelihood in topic models is an active research area [16, 17] due to the intractable sum over correlated $\mathbf{y}$ in Eq. (5). We take the view of [16, 17] and define an importance sampling approximation

$$p(\mathbf{x}^*|c) \approx \frac{1}{S} \sum_s \frac{p(\mathbf{x}^*, \mathbf{y}^s|c)}{q(\mathbf{y}^s|c)}, \ \mathbf{y}_s \sim q(\mathbf{y}|c), \tag{6}$$

where we drop conditioning on the parameters for clarity. Different choices of proposal $q(\mathbf{y}|c)$ induce different estimation algorithms. The optimal proposal $q_o(\mathbf{y}|c)$ is proportional to $p(\mathbf{x}^*, \mathbf{y}|c)$, which is unknown. We can however apply the variational mean field approximation $q_{mf}(\mathbf{y}|c) = \prod_i q_i(y_i|c)$ with minimal Kullback-Leibler divergence to the optimal proposal by iterating

$$q_i(y_i|c) \propto \left(\alpha_y^c + \sum_{l \neq i} q_l(y_l|c) \right) \hat{\Phi}_{x_i, y_i}. \tag{7}$$

Note that this importance sampling proposal (Eq. (7)) is much faster and more accurate than the standard approach [18, 16] of using posterior Gibbs samples in Eq. (6), which results in the unstable harmonic mean approximation to the likelihood – an estimator which has huge variance in theory and in practice [16]. This is important for us because classification speed and accuracy depends on the speed and accuracy of computing marginal likelihood (Eq. (5)).

In summary, to classify a new clip, we use the importance sampling approach defined in Eqs. (6) and (7) to compute the marginal likelihood (Eq. (5)) for each behaviour c (i.e. typical 0, rare 1,2,...) and hence the class posterior for that clip (Eq. (4)). This works because of the switching structure of the model. If a clip contains some known rare behaviour, only the sampler for that category can allocate the corresponding rare topics and obtain a high likelihood, while samplers for other categories will only be able to explain typical activities in the clip, thereby obtaining lower likelihood.

2.5 Synthesising Topics: Detection of Unknown Rare Behaviours

So far, the proposed MC-ΔLDA model is designed for classifying a new unseen clip as one of the known rare behaviour (or typical) classes. In practice, we also want to detect other types of rarely occurring behaviours that could interest users, but for which no examples were available during training. A standard strategy for detecting unknown rare behaviours is to detect statistical deviation from the learned model of existing behaviours. Using MC-ΔLDA, this corresponds to computing and thresholding the marginal likelihood of the entire MC-ΔLDA model $p(\mathbf{x}^*) = \sum_{c=0}^{N_c} p(\mathbf{x}^*|c)p(c)$ where $p(\mathbf{x}^*|c)$ is computed in Eq. (5). However, this equates to the standard approach of unsupervised behaviour detection and suffers from the same drawbacks as unsupervised topic models. To address this problem, we convert the problem of unknown rare behaviour detection to a MC-ΔLDA classification problem by utilising prior knowledge.

Our solution lies in employing existing learned topics to synthesise new topics which are likely to represent interesting (and plausible) unknown rare behaviours in a specific scene, and inserting them as new categories in MC-ΔLDA. To that end, we exploit prior knowledge, e.g. analysis of how known rare topics relate to typical topics. In particular, we explore prior domain knowledge about spatio-temporal conflict of multiple existing topics to synthesise new topics of unknown behaviours. Although these behaviours do not cover every possible case in the real world, they are usually of particular interest but can easily be missed by standard approaches based on outlier detection [1, 3–5]. The synthesising procedure is as follows:

1. **Detecting conflict regions**: For an existing topic ϕ_t, identify the spatial location of dominant words to generate a binary matrix $M(\phi_t)$ in the image domain. Compute and threshold the matrix $M = \sum_{t=1}^{N_t} M(\phi_t)$ and connect components to identify $N_{t'}$ potential conflict regions $\{R_{t'}\}$ in each of which N_r existing topics are spatially overlapped. Each region $R_{t'}$ is associated with a map $M_{t'}$ in image domain to indicate locations of the region.
2. **Synthesising new topics**: Given a map $M_{t'}$, compute $\phi_{t'} = \sum_{t=1}^{N_r} \phi_t$ in which N_r existing topics $\{\phi_t\}$ in Φ sharing space with the map $M_{t'}$. Identify dominant words in $\phi_{t'}$ and normalise their probabilities. Set the values of non-dominant words to $\frac{h}{\#\phi_{t'}}$, where h is the mean probability of dominant words in all existing topics Φ and $\#\phi_{t'}$ is the number of non-dominant words in $\phi_{t'}$. Normalise $\phi_{t'}$ to obtain a multinomial distribution.

Given the synthesised additional topics that represent plausible conflicts among existing topics, we employ the MC-ΔLDA framework for classifying unseen behaviours. This is performed as in Section 2.4 after concatenating the synthesised topics $\Phi_e = \{\phi_{t'}\}$ with existing topics Φ and adding another class label c' to represent all unknown behaviours.

3 Experiments

3.1 Learning Rare Topics in Simulated Data

To clearly understand our model's performance, we first evaluate our method using simulated data where ground truth is known (Fig. 3). We defined 11 topics (2D structural patterns), of which 8 (bar patterns) were typical and 3 (star patterns) were rare topics (Fig. 3(a)). For training, we generated (1) 500 documents (images) of only typical topics and (2) 10 documents for each rare category by sampling from the corresponding α^cs (Fig. 3(b)). Note that purely typical documents are visually very similar to those also containing rare topics, as both are dominated by the same typical topics which also share words with the rare topics. Topics Φ learned by our model using this dataset (Fig. 3(b)) are shown in Fig. 3(c). Clearly, despite the extreme conditions of few rare documents (2%), and dominant typical topics within each rare document, MC-ΔLDA is capable

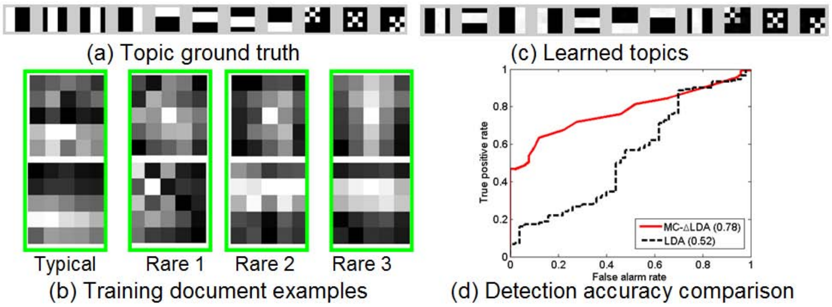


Fig. 3. Learning MC-ΔLDA to classify simulated data.

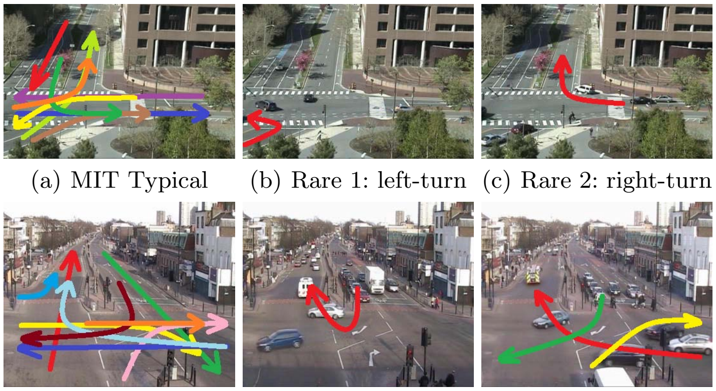
of learning correctly all the rare patterns in the training data. We then compared the performance of MC-ΔLDA against that of standard unsupervised LDA [18] for detecting rare topics in additional new testing documents generated separately from the training data. Standard LDA was trained using only typical documents and tested as an outlier detector. By varying the MC-ΔLDA priors $p(c|\varepsilon)$, the performance of both models can be compared using Receiver Operating Characteristic (ROC) curves and Area Under ROC values (in brackets), as shown in Fig. 3 (d). These results show a significant advantage of MC-ΔLDA over unsupervised LDA for detecting rare topics.

3.2 Learning Rare Behaviours in Public Space CCTV Videos

We validated the proposed framework using two road traffic video datasets: the MIT dataset [3] (30Hz, 720×480 pixels, 1.5 hour) and the QMUL dataset [4] (25Hz, 360×288 pixels, 1 hour). Both scenes featured a large number of objects exhibiting complex behaviours concurrently (see Fig. 4 for typical behaviours). To obtain the visual word codebook, the videos were spatially quantised into 72×48 cells and 72×57 cells respectively, and motion directions within each cell quantised into 4 orientations (up, down, left, right) resulting in codebooks of 13824 and 16416 words respectively. Each dataset was temporally segmented into non-overlapping short video clips of 300 frames each.

In each dataset, a number of clips containing two types of rare behaviours were manually identified and labelled into rare categories. The numbers of samples for each behaviour category are detailed in Table. 1. In the MIT dataset, these are vehicles turning left from below and turning right from the side (Fig. 4 (b) and (c), red arrows). In the QMUL dataset (Fig. 4 (e) and (f), red arrows), the rare behaviours are U-turns at the centre of the scene, and short-cut driving pattern (dangerous) in which vehicles drive into the junction before completion of normal driving patterns (yellow and green arrows), risking a collision.

Learning to Model Known Rare Behaviours – We first employed MC-ΔLDA to learn rare behaviours in the MIT dataset. Here we used 20 topics to represent typical behaviours and 1 topic for each type of rare behaviour. Fig. 5 (a) and (b) illustrate the dominant words in the learned topics for rare behaviours 'left-turn' and 'right-turn'. The learned topics accurately describe the known



(a) MIT Typical (b) Rare 1: left-turn (c) Rare 2: right-turn

(d) QMUL Typical (e) Rare 1: U-turn (f) Rare 2: collision

Fig. 4. Typical and rare behaviours in the experimental datasets.

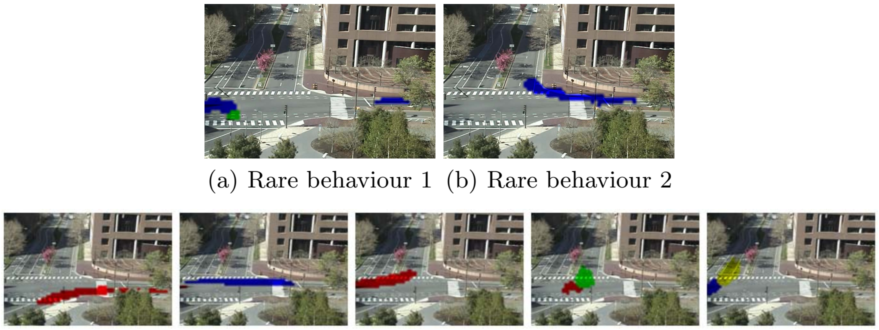
Table 1. Number of clips used in the experiments.

	MIT dataset			QMUL dataset		
	Typical (400)	Rare 1 (26)	Rare 2 (28)	Typical (200)	Rare 1 (12)	Rare 2 (5)
Training	200	10	10	100	4	2
Testing	200	16	18	100	8	3

rare behaviours illustrated by the examples in Fig. 4 (b) and (c). This is despite (1) variations in the spatial and temporal positioning of the rare activities within each training clip, i.e. the rare class generalisation is good; and (2) overwhelming co-occurring typical behaviours and heavily shared visual words between typical and rare behaviours (Fig. 5 (c)), i.e. the model was able to learn rare behaviours despite very biased training data.

Discovering rare topics in the QMUL dataset is harder because fewer training samples are available (e.g. 2 for Rare 2). Also the objects were highly occluded and motion patterns were frequently broken. For example, to complete a U-turn, vehicles usually drive to the central area and wait for a break in oncoming traffic before continuing. Employing one topic for each rare behaviour is therefore insufficient, and we applied two topics for each in this dataset. The results are shown in Fig. 6. Interestingly, rare topics are composed mostly of words from existing typical topics, but co-occurring in an unusual and unique combination. These uniquely composed rare topics will facilitate rare clip detection, because the correct rare model ($c > 0$) will be able to explain the rare behaviour with the use of a single rare behaviour (up to two topics), whereas the normal model ($c = 0$) will have to use many normal topics to explain the rare behaviour, thus obtaining lower likelihood (Eq. (5)).

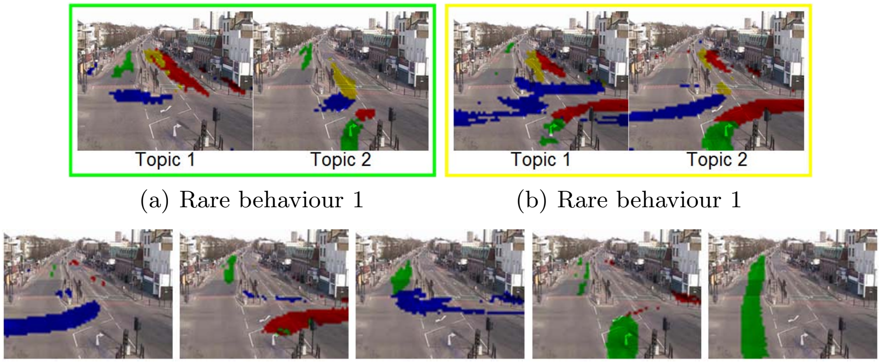
Detecting and Classifying Known Rare Behaviours – We compared the performance of our MC-ΔLDA model on classifying known rare behaviours in unseen data with that of an alternative LDA classifier [8] (LDA-C), in which independent LDA models were learned using clips from each class of training data (see Table 1). For LDA-C, we used 8 topics for each class of clips making



(a) Rare behaviour 1 (b) Rare behaviour 2

(c) Typical behaviours

Fig. 5. Learned MC-ΔLDA topics in the MIT dataset for the rare behaviours of interest (a) and (b), and example topics of typical behaviours (c). Colour indicates motion direction: red →, green ↑, blue ←, yellow ↓.

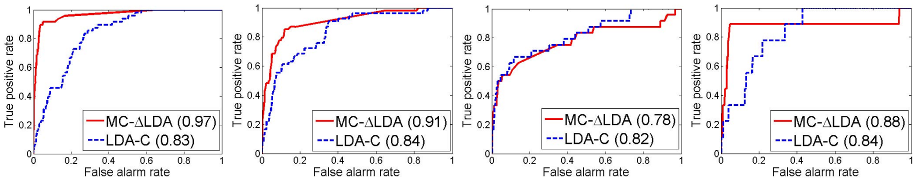


(a) Rare behaviour 1 (b) Rare behaviour 1

(c) Typical behaviours

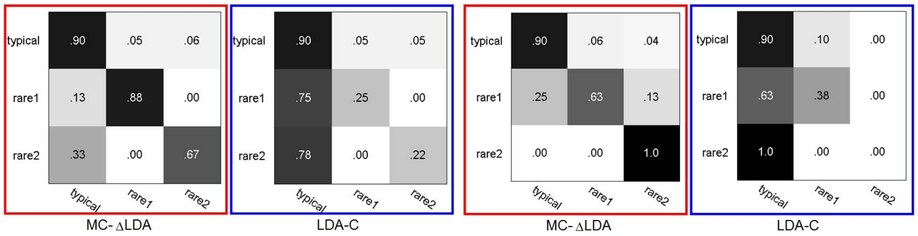
Fig. 6. Learned MC-ΔLDA topics in the QMUL dataset for the rare behaviours of interest. Colour indicates motion direction: red →, green ↑, blue ←, yellow ↓.

the total number of learned topics (24 topics) nearly the same as our MC-ΔLDA model. Classification was performed using the class posterior obtained from the individual LDA model likelihoods and corresponding priors. We first compared the performance of separating clips in one specific class from the others (i.e. detection). We varied the prior $p(c|\varepsilon)$ of the target class from 0 to 1 to produce ROC curves and Area Under ROC values. We randomly selected training and testing datasets in three folds and averaged the ROC curves. The results shown in Fig. 7 illustrate clearly the superior performance of MC-ΔLDA over LDA-C. While ROC curves are useful for evaluating detection performance, the simultaneous classification performance among all classes is not measured, which is important as perfect classification accuracy of one specific class could be at the expense of significant mis-classification among other classes. We thus also evaluated full classification performance, using confusion matrices. Note that setting different values of class prior $p(c|\varepsilon)$ determines classification performance for both models. In practice, low false alarm rate is critical for an automated rare



(a) MIT rare 1 (b) MIT rare 2 (c) QMUL rare 1 (d) QMUL rare 2

Fig. 7. ROC curves for identifying rare behaviours in unseen data.



(a) MIT dataset (b) QMUL dataset

Fig. 8. Classification comparison given a 10% false alarm rate.

behaviour detection system and it is a common criterion for setting parameters. We therefore set the priors to values that give a 10% false alarm rate for both models, and the classification performance is compared in Fig. 8. It can be seen clearly that given a low false alarm rate, our MC-ΔLDA yields a superior classification rate for the rare behaviours, with many fewer rare clips mistakenly classified as typical, leading to much higher true positive rate (77% vs. 23% and 73% vs. 28% for MIT and QMUL datasets respectively).

The inaccurate classification using LDA-C is the result of learning separate LDA models from different classes of corpa without sharing visual features. As a result, topics for the same type of behaviours would be represented differently (see the last two columns in Fig. 9). This duplicated representation of topics introduced noise to the classification. In contrast, this source of noise was removed in our MC-ΔLDA model, which jointly learned the exactly same topic representation for typical behaviours across all clips. Furthermore, as shown in Fig. 9, among the 8 learned topics there are topics that correspond to rare behaviour. However, a key difference between LDA-C and our MC-ΔLDA is that the former is not a multi-instance learning method – it solely learns a bag/clip classifier without automatically learning exactly which topics correspond to rare behaviours and thus should be relied upon for detection. In contrast, association between rare topics and rare behaviours was automatically learned using MC-ΔLDA and therefore this resulted in superior classification performance.

We also compared the performance of MC-ΔLDA with unsupervised LDA for behaviour detection by treating clips with any type of rare behaviour as outliers. To do so, an unsupervised LDA model with 20 topics was learned using only typical behaviour clips and the normalised likelihood of an unseen clip was used for detection. The results are shown in Fig. 10. Again MC-ΔLDA significantly outperformed LDA in both datasets. Unsupervised LDA was better at detecting

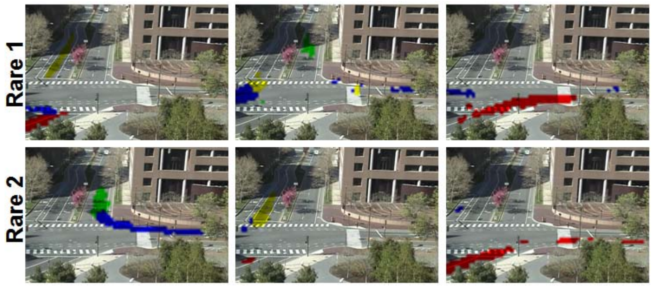
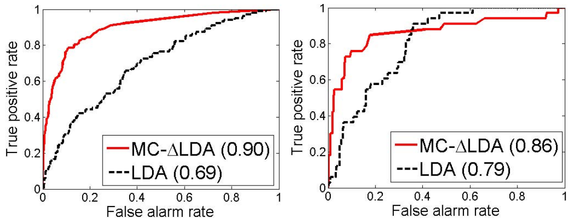


Fig. 9. Example topics learned using LDA-C on MIT dataset. Topics in the first column correspond to rare behaviours and the others correspond to typical behaviours.

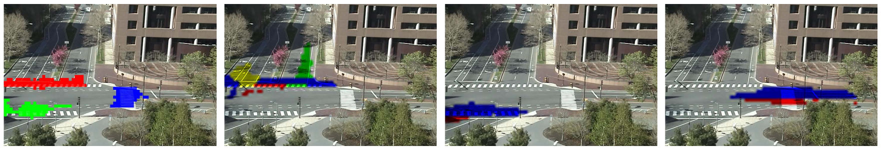


(a) MIT dataset (b) QMUL dataset

Fig. 10. Comparing MC-ΔLDA with unsupervised LDA for rare behaviour detection.

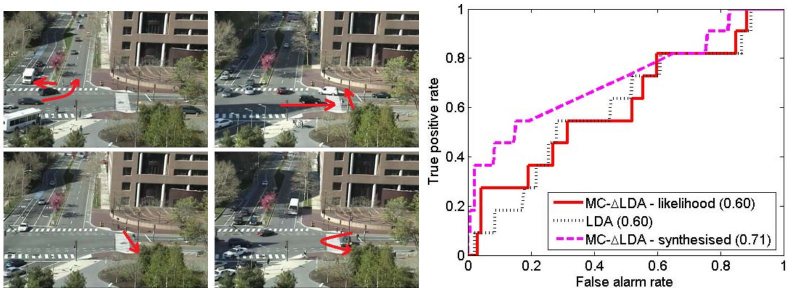
rare behaviours in the QMUL scene than in the MIT scene. This is because that, although cluttered, the rare behaviours in the QMUL dataset were caused by irregular co-occurrences of visual words from multiple objects which is exactly what LDA is designed for. The MIT scene is harder for LDA because visual features of rare behaviours were very subtle and overwhelmed by surrounding typical topics. In contrast, MC-ΔLDA reliably learned a precise description of all rare behaviours in both scenes making them easily detectable.

Detecting Unknown Rare Behaviours – Finally we conducted an experiment to illustrate the effectiveness of the MC-ΔLDA model on detecting unknown rare behaviours by synthesising topics using prior contextual knowledge. Using the method described in Sec. 2.5, three confict regions in the MIT scene with the potential for collision were automatically identified (Fig. 11 (a)). The three synthesised topics generated by our model are shown in Fig. 11 (b)-(d) respectively. Interestingly, these synthesised topics are in fact highly plausible showing the potential collisions either between vehicles and pedestrians or among vehicles. We used 11 clips containing different types of unknown interesting behaviours, including 'U-turn', 'Jay-walking', 'Vehicle-pedestrian near collision' to evaluate the performance (see Fig. 12 (a)). MC-ΔLDA with synthesised topics was compared against two detection methods: (1) the normalised marginal likelihood of MC-ΔLDA over all three known classes; (2) unsupervised LDA using normalised likelihood. Our results in Fig. 12 (b) show significant improvement by the proposed new method. Indeed, using MC-ΔLDA likelihoods for detection bears the exactly same drawback as unsupervised LDA – of being less



(a) Conflict regions (b) Syn. topic 1 (c) Syn. topic 2 (d) Syn. topic 3

Fig. 11. Conflict regions and synthesised topics.



(a) Examples of unknown behaviours (b) ROC curves for detection

Fig. 12. Comparison for detecting unknown behaviours in unseen data.

sensitive to subtle behaviours without distinctive visual features. In contrast, MC-$\varDelta$LDA with synthesised topics incorporating prior contextual knowledge about risky/interesting potential local behaviour conflicts is more sensitive to less obvious unknown rare behaviours.

4 Conclusion

We introduced Multi-Class Delta Latent Dirichlet Allocation (MC-$\varDelta$LDA), a novel approach for learning to detect and classify rare behaviours in video with weak supervision. By constraining a topic model prior depending on a video clip class, we are able to learn explicit models of rare behaviours jointly with "background" typical behaviours. Subsequently, by using importance sampling to integrate out the topics of a new clip under this prior, we can accurately determine the behaviour class of an unseen clip. This is $\mathcal{O}(SNN_c)$ for S samples, N words and N_c behaviour classes, which was real-time for the MIT dataset.

Standard unsupervised outlier detection methods [3–5] often fail to detect rare behaviours because each clip contains mostly irrelevant typical features. In contrast to those approaches, the proposed method has two further advantages: the ability to *classify* rare behaviours into different classes, and the ability to *locate* the offending sub-regions of the scene. With conventional supervised classification, lack of rare class data prevents good model learning. Our approach, which can be considered as a supervised multi-instance learning generalisation of LDA, mitigates these problems effectively by performing joint localisation and detection during learning. The data are used efficiently, as typical behaviour features are shared between all classes, and the few rare class parameters in the model focus on learning the unique features of the associated rare classes. Compared to LDA-C, this sharing of typical features improves performance by eliminating a source of classification noise (different typical representation), and ensuring each rare behaviour is well modeled (via the constrained prior), even in the extreme case of only a single rare class training example.

Important areas for future research include generalising MC-ΔLDA to online adaptive learning for adapting to changing data statistics, and active learning where clip labels can be incrementally and intelligently queried to discover new behaviour classes of interest and learn to classify them. Finally, we proposed a novel although partial solution to the issue of detecting new/unknown rare behaviours, using very general prior knowledge about the scene to synthesise new topics that predict new behaviours (Sec. 2.5). A major area of future research is to formulate a more general model for transfer learning of existing rare behaviour topics to predict new and unknown rare behaviour topics.

References

1. Xiang, T., Gong, S.: Video behavior profiling for anomaly detection. PAMI **30** (2008) 893–908
2. Saleemi, I., Shafique, K., Shah, M.: Probabilistic modeling of scene dynamics for applications in visual surveillance. PAMI **31** (2009) 1472–1485
3. Wang, X., Ma, X., Grimson, E.: Unsupervised activity perception by hierarchical bayesian models. PAMI **31** (2009) 539 – 555
4. Li, J., Gong, S., Xiang, T.: Global behaviour inference using probabilistic latent semantic analysis. In: BMVC. (2008)
5. Hospedales, T., Gong, S., Xiang, T.: A markov clustering topic model for behaviour mining in video. In: ICCV. (2009)
6. Xiang, T., Gong, S.: Beyond tracking: Modelling activity and understanding behaviour. IJCV **61** (2006) 21–51
7. Robertson, N., Reid, I.: A general method for human activity recognition in video. Computer Vision and Image Understanding **104** (2006) 232–248
8. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of Machine Learning Research **3** (2003) 993–1022
9. Viola, P., Platt, J., Zhang, C.: Multiple instance boosting for object detection. In: NIPS. (2005)
10. Nguyen, M.H., Torresani, L., de la Torre, F., Rother, C.: Weakly supervised discriminative localization and classication: a joint learning process. In: ICCV. (2009)
11. Blei, D., McAuliffe, J.: Supervised topic models. In: NIPS. Volume 21. (2007)
12. Andrzejewski, D., Mulhern, A., Liblit, B., Zhu, X.: Statistical debugging using latent topic models. In: ECML. (2007)
13. Markou, M., Singh, S.: A neural network-based novelty detector for image sequence analysis. PAMI **28** (2006) 1664–1677
14. Abe, N., Zadrozny, B., Langford, J.: Outlier detection by active learning. In: KDD. (2006) 504–509
15. Minka, T.P.: Estimating a dirichlet distribution. Technical report, Microsoft (2000)
16. Wallach, H., Murray, I., Salakhutdinov, R., Mimno, D.: Evaluation methods for topic models. In: ICML. (2009)
17. Buntine, W.: Estimating likelihoods for topic models. In: ACML. (2009)
18. Griffiths, T.L., Steyvers, M.: Finding scientific topics. Proceedings of the National Academy of Sciences **101** (2004) 5228–5235