Video Topic Modelling with Behavioural Segmentation

Tom S. F. Haines & Tao Xiang
{thaines,txiang}@dcs.qmul.ac.uk

Queen Mary, University of London

29th October 2010

Roadmap

Note that code can be obtained from *thaines.net*

Probabilistic Models & Abnormalities

- Goal is to detect abnormal behaviour.
- Have lots of normal data, but little if any abnormal data.

$$\therefore$$

- Build (generative) probabilistic model of normal behaviour.
- Detect abnormalities as unlikely data, i.e. outliers.

Video Features

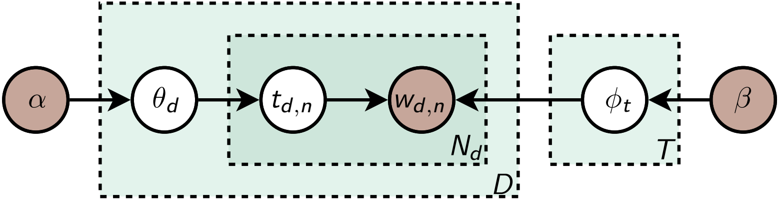
- Need to extract features from the video.
- Use quantised optical flow - sampled on a grid with the four compass directions, with thresholds.
- Video is split into short clips, each of which contains a bag of these features.

Topic Models

- Model originally proposed for corpora of documents that contain words.
- Learns topics - distributions over words shared between multiple documents.
- Ignores order of words - time and space are not considered.
- For video data clips are mapped to documents and features to words.

LDA

- *Latent Dirichlet Allocation* is one specific, and very successful, topic model.
- Proposed algorithm is an extension.



Issue 1: Sensitivity

- CCTV typically captures large areas, and large numbers of actors.
- Most actors will be behaving normal, whilst only a small number, typically one, will not.
- The deflection in the probability caused by the abnormal actor(s) is often of comparable magnitude to the noise.

Issue 2: Localisation

- These systems are unreliable - a human needs to verify each detected issue.
- But the system provides a single score for each clip, and in a multi-camera configuration, a single score for *all* cameras.
- A human would therefore have to search for the abnormality in all the data, which is not reasonable.

Behavioural Segmentation

- Different behaviours occur in different regions of a scene - the scene can therefore be segmented by the behaviours.
- Many authors have proposed stand-alone solutions, often based on clustering.
- This work proposes a topic model that includes a behavioural segmentation, with the entire model learned simultaneously.

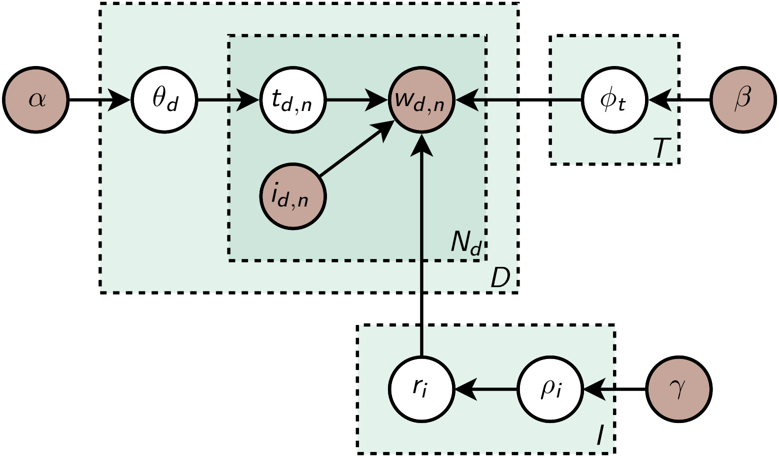
Relation to Sensitivity & Localisation

- Abnormal clips are recognised as having a low probability of being drawn from the model - one value for the entire clip...
- ...with regions the probability can be calculated for each region.
- *Sensitivity* is improved by looking for regions in each clip with a low probability.
- *Localisation* is improved by highlighting the specific region that has a low probability.

Concept of Regions

- In LDA the location and direction of motion for a feature are combined for each word.
- For the presented location (Identifier, i) and the direction of motion (Word, w) remain separate.
- A topic now defines a distribution over the direction of motion and *region*, r, for each sample drawn from it.
- Each location belongs to a specific region.
- Consequentially locations are clustered into regions where topics have similar distributions over words.

Graphical Model



Gibbs Sampling

- Gibbs sampling is used to sample model parameters.
- Consists of two steps - the *t-step* and the *r-step*.
- The t-step re-samples the $t_{d,n}$ values, and is the obvious adjustment of Griffiths and Steyvers '04.
- The r-step re-samples the region assignment for each identifier.

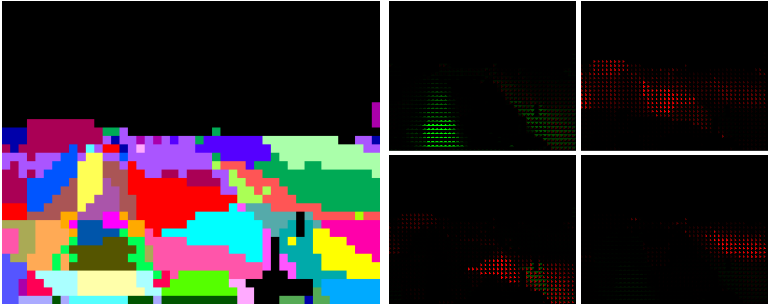
Post-processing

- Each location is given a hard assignment to a region for each sample drawn, but the model contains ϕ, which we want to estimate, a task which requires many samples.
- Therefore multiple Gibbs samples of the model parameters have to be merged.
- Problem: Each sample can have different topic and region identifiers - need to match them up.
- Done with an algorithm based on greedy symmetric Kullback-Leibler divergence.
- As their is no guarantee that the same topics/regions exists between any two samples this can theoretically fail, but in practise this does not happen.

Mile End Dataset

- 50 minute long video sequence of a traffic junction.
- Has three phases.

Learnt Model

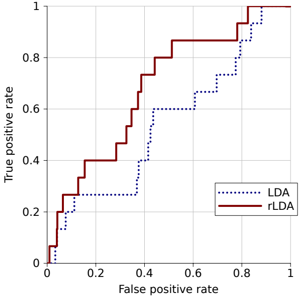


(a) Extracted regions (b) Example topics

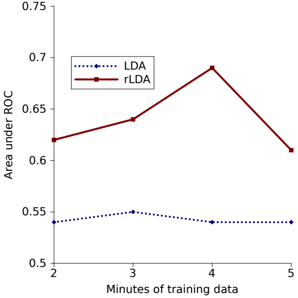
- Note: Regions are actually soft, and blend into each other - the figure shows the most probable region for each quantised location.

Abnormality Detection

- For each document take the five least probable regions and multiply their probabilities. Abnormal documents are then defined as those with a probability lower than a threshold.
- *Receiver operating characteristic* [ROC] curves may be generated:



(c) ROC with 4 minutes of training data



(d) Area under ROC for varying amounts of training data

Localisation Successes

(e) Ambulance (bottom right).

(f) U-turn (top left).

(g) Car cuts in front (top right).

(h) Person crossing in wrong place plus a u-turn (top left).

(i) Run across road (middle left).

(j) Person crossing during traffic, including a u-turn (top left).

Localisation Failures

(k) Police car (middle).

(l) Temporal quantisation issue.

(m) Fireengine (middle).

(n) Cement truck (middle).

(o) U turn (top middle).

(p) Confused region (top left).

More Localisation Examples

(q) Birds landing in road (top left).

(r) Bike threading between cars (bottom middle).

(s) Car uses gap in traffic (middle).

(t) Pedestrian and car use crossing at same time (top left).

(u) Electric scooter takes different route (top left).

(v) Pedestrians and cars (top left).

Conclusions

- Demonstrated *region* LDA - LDA with a behaviour segmentation.
- Joint learning of a behavioural segmentation with a topic model assist with abnormality detection.
- The segmentation gives improved results with regards to both the sensitivity and localisation issues.
- This is an idea in progress - the behavioural regions implicitly defined by this algorithm can almost certainly be improved upon.

The End

Questions?