DELTA-DUAL HIERARCHICAL DIRICHLET PROCESSES: A PRAGMATIC ABNORMAL BEHAVIOUR DETECTOR

EPSRC Engineering and Physical Sciences Research Council

TOM S. F. HAINES & TAO XIANG

{thaines,txiang}@eecs.qmul.ac.uk

Queen Mary University of London

ABSTRACT

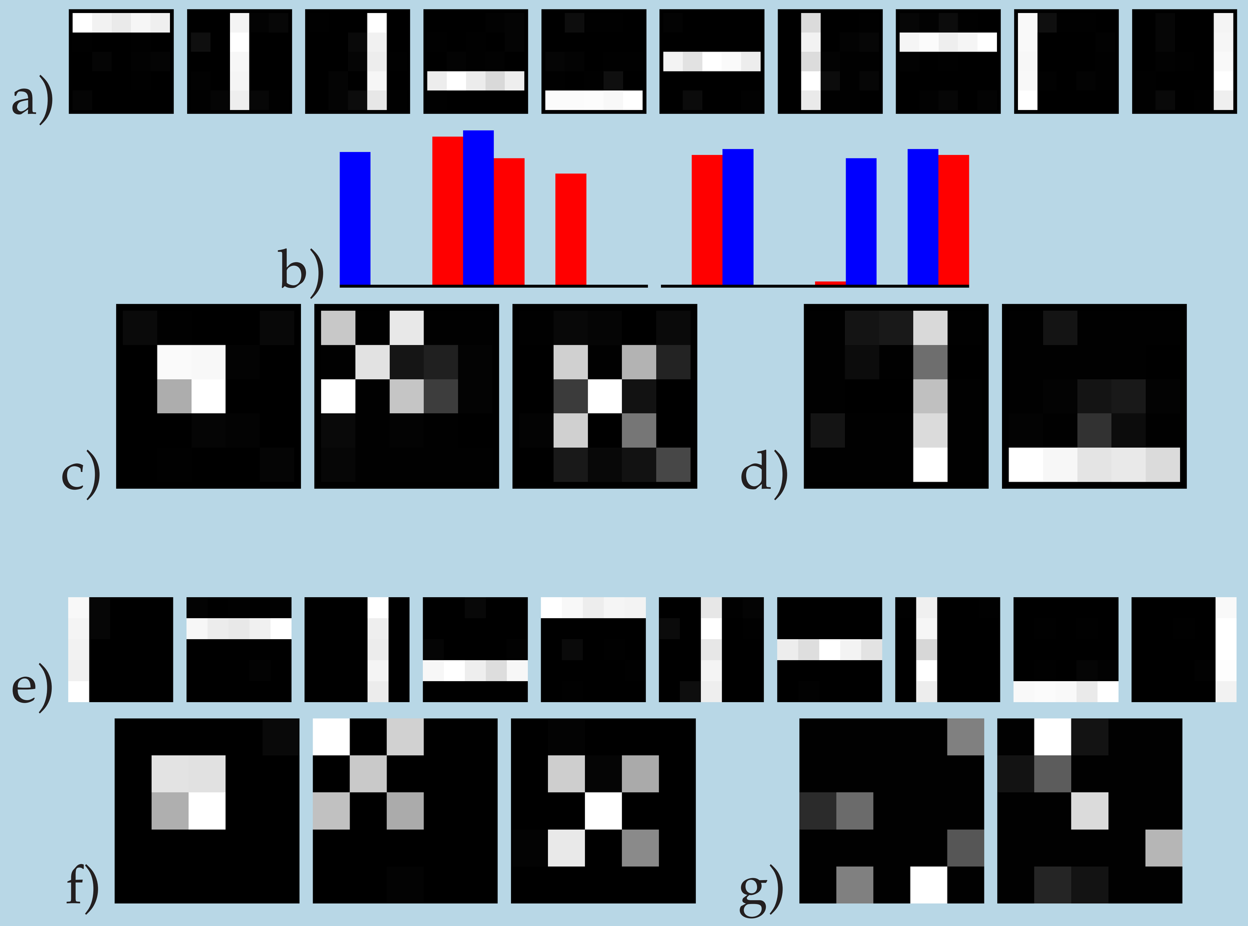
In the security domain a key problem is identifying rare behaviours of interest. Training examples for these behaviours may or may not exist, and if they do exist there will be few examples, quite probably one. We present a novel weakly supervised algorithm that can detect behaviours that either have never before been seen or for which there are few examples. Global context is modelled, allowing the detection of abnormal behaviours that in isolation appear normal.

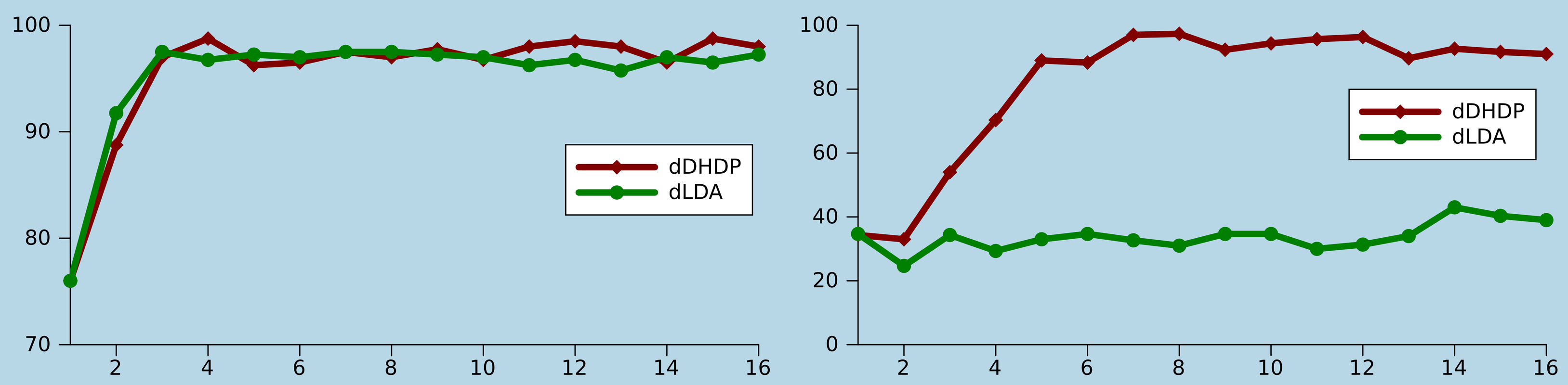
SOLUTION OVERVIEW

- Topic models such as Latent Dirichlet Allocation [LDA] can model the behaviour in a video. Features [words] are extracted from the video, and then the video is split into short clips [documents]. Each clip is modelled as a mixture of simultaneously occurring behaviours [topics].
- More sophisticated methods such as Dual Hierarchical Dirichlet Processes [DHDP] cluster the video clips, and hence model the normal co-occurrence of behaviours. They also use non-parametric Bayesian methods, which saves on parameter tuning.
- LDA and its relatives are unsupervised models, and abnormal behaviour can only be detected by its low probability.
- delta LDA [dLDA] allows for semi-supervised learning, such that if you have a video clip with abnormal behaviour in it you can train the topic model to classify that behaviour. Unsupervised detection continues to work.
- We introduce Delta-Dual Hierarchical Dirichlet Processes [dDHDP], which combines these two methods, and gains their respective advantages. It has the novel ability to learn models for abnormal behaviours that consist of normal behaviours occurring in an abnormal context.
- As an example a person normally crosses the road when there is no traffic - this fact is learned, and a person crossing the road during traffic will appear abnormal. A model of this behaviour can be learnt given few training examples, possibly only one, allowing its future detection.

DEMONSTRATION

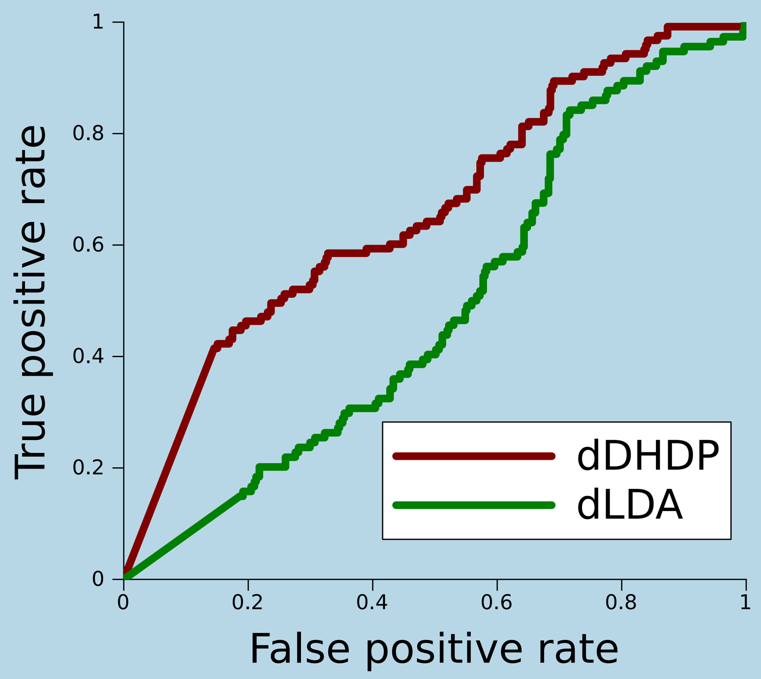
- This demo uses a set of documents with a $5X5$ grid of words, visualised as pixels in an image.
- There are 10 topics - 5 vertical and 5 horizontal lines. Only one orientation is in each document.
- There are 3 abnormal topics and 2 cases of normal topics appearing in the wrong document.
- *a-d* are the results for dDHDP; *e-g* the results for dLDA. The normal topics (*a* and *e*) and abnormal topics (*c* and *f*) are learned by both approaches.
- dDHDP infers the document clusters , as given by the histograms *b*.
- dDHDP discovers normal topics at abnormal times (*d*), a task at which dLDA fails (*g*).





The graphs show performance as the number of training examples increases. The left is for the abnormal documents, where dLDA succeeds, the right for normal topics in abnormal contexts, where dLDA does no better than chance.

RESULTS



- Results for the mile end data set are presented.
- Two kinds of abnormality are used - a u-turn (above, left) and driving from the middle area to the right whilst traffic continues to travel vertically (above, right).
- Confusion matrices are presented to the right, demonstrating that dDHDP is the better approach.

- It is possible to do dual discovery - both supervised and unsupervised learning.
- Using a set of abnormal behaviours on which the algorithm is *not* trained, in addition to the two behaviours on which it is trained, a ROC curve is generated (above, centre). dDHDP again shows an advantage.

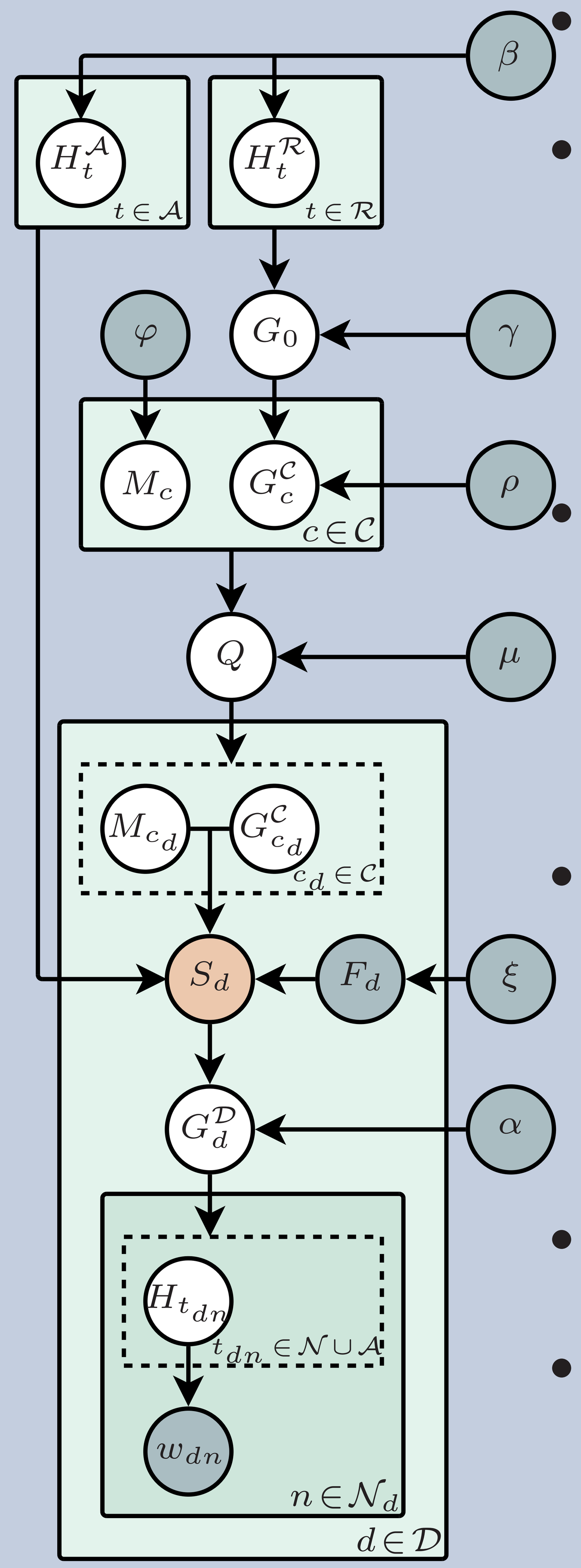
dDHDP = 83.7%

364	13	37	87.9%
4	6	1	54.5%
22	3	45	64.3%

dLDA = 74.2%

351	22	41	84.8%
0	11	0	100.0%
56	8	6	8.6%

SPECIFICS



- The graphical model is shown to the left.
- Both priors and non-parametric Bayesian methods are used - with sensible defaults for the priors no tweaking is needed.
- Base measures are represented by a box containing the entities drawn from them, connected to the relevant Dirichlet process [DP].
- S_d combines a DP with a multinomial, to create a set of user provided abnormal topics with the number of normal topics learned.
- A collapsed Gibbs sampler is used.
- Real time analysis of new clips is obtained with a varient of the left to right algorithm.

CONCLUSION

- The approach captures a class of behaviours that previous approaches could not...
- ...but it does so at the expense of a very complex model.
- It takes a long time to train the model...
- ...though can run in real time when analysing new documents.
- Simpler models should be tried, to see if they can do the same thing.
- Its hard to know when it has converged.

MORE...

In addition to the paper an implementation in python can be obtained from *thaines.com*.