

Passenger Abnormal Behaviour Detection using Machine Learning Approach

Sonal Dabral¹, Amit Agarwal², Sachin Kumar³, Bhaskar Gautam⁴

¹Computer Science & Engineering Department, Dehradun, Graphic Era University,

²Computer Science & Engineering Department, Indian Institute of Technology Roorkee

³Centre for Transportation Systems, Indian Institute of Technology Roorkee

⁴Computer Science & Engineering Department, National Institute of Technology Karnataka

Abstract—In this paper we have proposed the clustering approach to classify the random walk trajectories from the synthetic bus station video. Bus station one of the most crowded locations that consist of more than thousands of passengers or travelers waiting for the buses to travel to the destination point. These crowded locations can be highly prone to accidents or terrorist activities. Work is classified into two steps i.e Firstly we find out the trajectories from the image by using the machine learning approach after that we apply the agglomerative clustering algorithm which is used to group the abnormal trajectories with the similar spatial patterns and normal trajectories with similar spatial patterns.

Keywords—Path detection, Anomaly Detection, Trajectories, Clustering.

I. INTRODUCTION

Data mining is the extraction of intelligent information from a large set of data [1-2]. This information can be utilized to provide different solutions for real time problems. Data mining has a proven track record in almost every domain such as bioinformatics, healthcare data analysis, transportation, social media analysis etc [3-14]. Popular data mining techniques are classification, clustering, association rule mining and anomaly detection [15]. Anomalies are data objects with abnormal behavior. Anomaly detection can be defined as a sub-domain of data mining that deals with identification of abnormal data objects. Bus stations or bus depot can be defined as a crowded location that consists of more than hundreds of passengers or travelers waiting for the buses to travel to the destination point. These crowded locations can also be a point that can be highly prone to accidents or terrorist activities. The persons involved in these bad activities have some kind of abnormal behaviors due to some psychological effect on their mind. Therefore these persons can be treated as anomalies in the crowd of passengers. If these anomalies can be detected due to some means, it can help in preventing bad activities in the bus depot or bus stations. The surveillance camera is a great source of data in the form of images captured at bus stations. This image data can be analyzed to perform analysis using anomaly

detection techniques and further persons with anomalous behavior can be detected for further investigation. Various studies on anomaly detection using image data exist in the literature [16-25]. In this study, security issues and maintenance of a bus station is the primary focus. The anomaly detection at bus stop surveillance data is very important as it can lead to detection of accidents or mischievous behavior by people in real time. Therefore, it can be utilized to detect and alert any kind of anomalous behavior in real time. The assumption made for this study is based on the fact that the images are captured through a surveillance device at the bus station. These images are then used for processing using data mining techniques to detect any anomalous behavior of any individual who may be involved in wrong activities. The automation of this process would be helpful to reduce the human efforts and to increase the security of the bus station. However, the detection of anomaly in the crowd is a hard problem that we trying to solve. The organization of the paper is as follows: Section 2 will present the methodology and data set used in this study. The results and discussion is presented in section 3 and a conclusion is provided in section 4.

II. MATERIALS AND METHODS

This section presents a description of the data and methodology used in this study.

A) Dataset generation

In this paper, as Data Mining is main emphasis, preprocessing and tracking parts of solution are not implemented as they belong mostly to image processing domain. Instead a dataset is generated to get a supposedly real dataset which simulates surveillance data. A 20x20 graph is taken to represent the bus station. Three blocks are taken as entrance blocks and three are taken as exits. Now the spatial- temporal paths of objects are generated in following way, for every combination of entrance and exits few intermediary nodes are selected arbitrarily and thus random paths are generated. These arbitrary paths thus generated represent the path followed by a

particular object's/individual. We consider the fact that the number of anomalous paths will be considerably less than the non anomalous paths while generating the dataset. We consider a path non-anomalous when the person/ objects enter and leave the block without any diversions/ delay. Also there are paths such that objects is entering but never exiting or exiting after following a very long path, these paths are anomalous and should be detected by the implemented algorithm.



Figure 1. Sample of Bus station video



Figure 2. Image Frame is extracted from video

The dataset is all the coordinates in each path. For example in this sample, the first column represents the x-coordinates and second column represents the y-coordinate. The third column is for the purpose of identifying where the next move of the random walker will be horizontal or vertical. In which 1 represent the random walker may move in the vertical direction were as 2 represents the random walker may walk in the horizontal directions.

Table 1. A part of dataset used

X-coordinates	y-coordinates	People move in vertical direction	People move in horizontal direction
34	21	0	2
33	21	1	0
33	22	1	0
33	23	0	2
32	23	1	0
32	23	0	2
34	42	0	2
42	33	0	2
32	23	1	0
44	37	1	0
38	32	1	0
33	22	0	2
49	44	1	0

B. Preprocessing

Data preprocessing [2] plays a very important role in data analysis. CCTV cameras are installed in most of the crowded region and which generate a lot of data in the form of frames but this data is not used directly in the system, so we have to first perform the preprocessing because there is high

frequency noise is generally critical to be identified and removed. So this has to be done by converting video frames in to gray scale images. After converting into gray scale images it should be filtered through Gaussian low- pass filter. Then apply the object detection technique of frame subtraction is applied.

Figure 3 is sample image which is taken from the video that is captured from bus station and image extraction process is already explained in the above paragraph, were as this image is shows the trajectories.

In the below figure 3 we have shown the spatial model that we have developed for representing the routes from the images. In which each path having number of nodes which are bounded by the envelope and which is used to determine the variation of the trajectories in the route where each route having end and exit point. And all the nodes are equal separation distance.

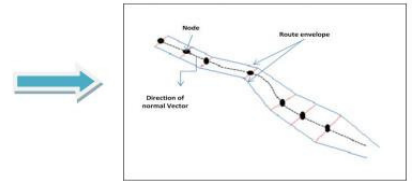


Figure.3 Sample trajectories are plotted in Spatial Model of a route.

C. Tracking

Once the objects are detected, these objects are tracked using various attribute like contrast, Correlation, Energy and homogeneity for each frame [21]. Several objects tracking algorithms can be used to track the objects and obtain the path followed by each object's/ person to decent accuracy. In the below figure 3 we have shown the spatial representation of route. Once we are having the path followed by each object's/person then we can use Data mining techniques to analyze these motions and detect and flag an anomaly if present.

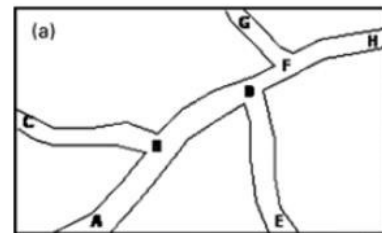


Figure 4. Spatial representation of route

D. Clustering

Once we have the paths of each object's/person our next job is to detect anomaly in these paths. For this our basic approach will be to use a clustering algorithm to cluster similar paths. All the outliers that emerge out as a result of the clustering algorithm can be flagged and called as anomaly. Explaining the clustering process in details, these paths are clustered using fixed width clustering methods so that similar paths are

in the same cluster. As the second step, similar clusters are merged based in the distance between clusters. If the average intercluster distance is more than φ number of standard deviations of the inter-cluster distance from the mean inter cluster distance, than a cluster is declared as anomalous. This anomaly detection algorithm is further explained in the pseudo code section.

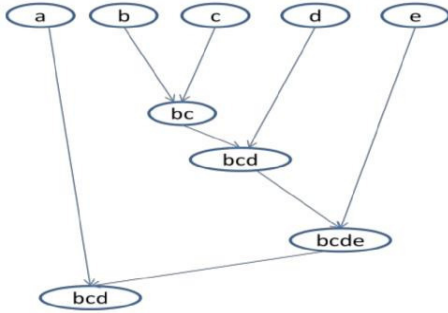


Figure.5 Agglomerative hierarchical clustering

Steps involved in agglomerative hierarchical clustering as explain below.

1. Calculating the proximity between the two trajectory using the weighted Euclidean distance and average linkage method and variable number of clusters.
2. Then find the most similar pairs of cluster by using the proximity matrix. Combine two patterns into one cluster and update the proximity matrix to reflect this merge operation.
3. Do above steps until all the patterns in one cluster. Maintaining the Integrity of the Specifications

E. Distance Measure for Clustering (DMC)

For clustering the paths and extracting meaningful clusters we need to choose a good distance measure. Given that we have many entry and exit points, there are many possibilities of non-anomalous paths. We need to choose a distance function as follows: Let P_1, P_2 be any two paths represented by array of 20 20 matrix where the paths is represented by joining of 1s.

Distance measure for clustering (DMC) is calculated by using the below formula.

$$D(P_1, P_2) = \sqrt{\left(\frac{w_1}{w_1 + w_2}\right)(l_1 - l_2)^2 + \left(\frac{w_2}{w_1 + w_2}\right)(\cos(\text{avg}P_1) - \cos(\text{avg}P_2))^2} \quad (1)$$

Pseudo Code

Algorithm 1: Anomaly Detection

1. Apply preprocessing step and identify foreground object.
2. Obtain the tracks/paths as explained using the *Kalman Filter*.
3. Obtain the length of path and average of cosine of all vectors present on the path to represent the path in the form of a point on 2D space.
4. Apply *agglomerative* clustering with distance metric as weighted Euclidean distance and average linkage method and variable number of clusters.
5. Identify the clusters corresponding to anomalous paths and report the corresponding paths.

III. RESULTS AND DISCUSSION

Firstly we have shown the number of trajectories that we have obtained from the above image. In which all the normal path is represented by the white line from the entry to exit gates. The blue line is anomalous path following in the different trajectory. For this we have use the kalman filtering in matlab.

Steps are shown below.

1. Firstly we create the vision. KalmanFilter by using the KalmanFilter.
2. Then to eliminate the noise present in the image we use predict and correct methods.
3. When the trajectory is detected, then its firstly predicts its state at the current video frame, then on the newly detected trajectory it uses previous detected object location to correct its state.

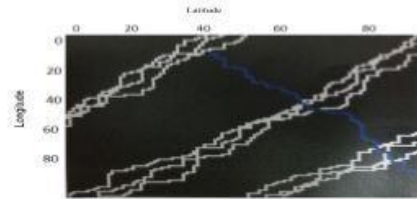


Figure 6 A plot of all trajectories in the dataset in which 10 random trajectories are highlighted in color.

After that we apply the agglomerative clustering algorithm, in which each data points combine as singleton cluster, until all the data points have been combining to form the single cluster.

In the below figure the clustering results are shown. The clustering results are as expected we get four separate clusters corresponding to a four different trajectories with clusters belonging to normal paths being close to each other and the one with anomalous path is segregated

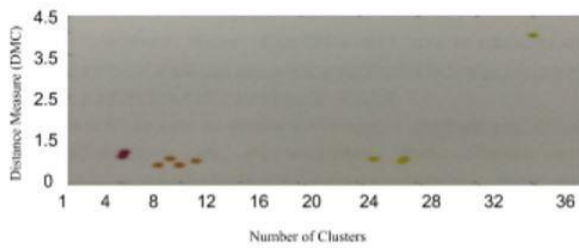


Fig.7 Results of agglomerative Clustering

We have used the agglomerative hierarchical clustering algorithm because it has many advantages as ease of handling of any type of similarity or distance, flexible in regarding a level of granularity and it can be applied in any type of attribute.

IV. CONCLUSION

This paper deals with a practical approach of anomaly detection from the bus station. For that we have practically analyzed the trajectory data that we have extracted from the image. Then finding the nodes in the path so we can easily identify whether the path is abnormal or normal. We applied the agglomerative clustering algorithm to obtain the abnormal clusters. In my knowledge, no one has before applied anomaly detection in the bus station although it is a crowded location that consists of more than hundreds of passengers or travelers waiting for the buses to travel to the destination point. These crowded locations are highly prone to accidents or terrorist activities. So in the future work we may apply this practical approach in the real dataset. Or it can be helpful in many applications like in anomaly detection in the airport etc.

References

1. Han J and Kamber M., Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, USA, 2001.
2. Tan PN, Steinbach M and Kumar V, Introduction to data mining. Pearson Addison-Wesley, 2006.
3. Kumar, Sachin, and Durga Toshniwal. "A data mining framework to analyze road accident data." *Journal of Big Data* 2, no. 1 (2015): 26.
4. Liu, Huiqing, Jinyan Li, and Limsoon Wong. "Use of extreme patient samples for outcome prediction from gene expression data." *Bioinformatics* 21, no. 16 (2005): 3377-3384.
5. Kumar, Sachin, and Durga Toshniwal. "Analysing road accident data using association rule mining." In *Computing, Communication and Security (ICCCS), 2015 International Conference on*, pp. 1-6. IEEE, 2015.
6. Hirschman, Lynette, Jong C. Park, Junichi Tsujii, Limsoon Wong, and Cathy H. Wu. "Accomplishments and challenges in literature data mining for biology." *Bioinformatics* 18, no. 12 (2002): 1553-1561.
7. Kumar, Sachin, and Durga Toshniwal. "A data mining approach to characterize road accident locations." *Journal of Modern Transportation* 24, no. 1 (2016): 62-72.
8. Aggarwal, Charu C. "An introduction to social network data analytics." In *Social network data analytics*, pp. 1-15. Springer US, 2011.
9. Kumar, Sachin, and Durga Toshniwal. "A novel framework to analyze road accident time series data." *Journal of Big Data* 3, no. 1 (2016): 1-11.
10. Getoor, Lise, Nir Friedman, Daphne Koller, and Benjamin Taskar. "Learning probabilistic models of relational structure." In *ICML*, vol. 1, pp. 170-177. 2001.
11. Kumar, Sachin, Durga Toshniwal, and Manoranjan Parida. "A comparative analysis of heterogeneity in road accident data using data mining techniques." *Evolving Systems* (2016): 1-9.
12. Pieter, Ben Taskar Ming-Fai Wong, and Abbeel Daphne Koller. "Link prediction in relational data." (2003).
13. Kumar, Sachin, and Durga Toshniwal. "Analysis of hourly road accident counts using hierarchical clustering and cophenetic correlation coefficient (CPCC)." *Journal of Big Data* 3, no. 1 (2016): 1-11.
14. Richard, R. J. A., and N. S. S. "A feasibility study of challenges and opportunities in computational biology: A Malaysian perspective." *American Journal of Applied Sciences* 2, no. 9 (2005): 1296-1300.
15. Chandala, V., A. Banerjee, and V. Kumar. "Anomaly Detection: A Survey, ACM Computing Surveys." *University of Minnesota* (2009).
16. Kim, Seong Soo, and AL Narasimha Reddy. "A study of analyzing network traffic as images in real-time." In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 3, pp. 2056-2067. IEEE, 2005.
17. Lakhina, Anukool, Mark Crovella, and Christophe Diot. "Mining anomalies using traffic feature distributions." In *ACM SIGCOMM Computer Communication Review*, vol. 35, no. 4, pp. 217-228. ACM, 2005.
18. Mirkovic, Jelena, and Peter Reiher. "A taxonomy of DDoS attack and DDoS defense mechanisms." *ACM SIGCOMM Computer Communication Review* 34, no. 2 (2004): 39-53.
19. Gyaourova, Aglika, Chandrika Kamath, and S. C. Cheung. "Block matching for object tracking." *Lawrence Livermore National Laboratory* (2003).
20. Shen, Ke, and Edward J. Delp. "A fast algorithm for video parsing using MPEG compressed sequences." In *Image Processing, 1995. Proceedings.*

- International Conference on*, vol. 2, pp. 252-255. IEEE, 1995..
21. Lienhart, Rainer, Christoph Kuhmunch, and Wolfgang Effelsberg. "On the detection and recognition of television commercials." In *Multimedia Computing and Systems' 97. Proceedings., IEEE International Conference on*, pp. 509-516. IEEE, 1997.
 22. Zhang, HongJiang, Atreyi Kankanhalli, and Stephen W. Smoliar. "Automatic partitioning of full-motion video." *Multimedia systems* 1, no. 1 (1993): 10-28.
 23. Kim, Hyogon, Inhye Kang, and Saewoong Bahk. "Real-time visualization of network attacks on high-speed links." *IEEE network* 18, no. 5 (2004): 30-39..
 24. Krishnamurthy, Balachander, Subhabrata Sen, Yin Zhang, and Yan Chen. "Sketch-based change detection: methods, evaluation, and applications." In *Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, pp. 234-247. ACM, 2003..
 25. Kim, Seong Soo, AL Narasimha Reddy, and Marina Vannucci. "Detecting traffic anomalies through aggregate analysis of packet header data." In *International Conference on Research in Networking*, pp. 1047-1059. Springer Berlin Heidelberg, 2004.