DISCOVERING THE TOP-K UNEXPLAINED SEQUENCESIN TIME-STAMPED OBSERVATION DATA

Bina Bhandari
Asst. Professor, Department of Comp. Sc. & Info. Tech., Graphic Era Hill University, Dehradun, Uttarakhand India 248002

Abstract
For instance, it may be necessary to spot odd occurrences in online transactions or in footage of an airport tarmac in a number of applications where it is necessary to recognise unusual behaviours in succession of time-stamped observation data. In this essay, we'll begin by focusing on a well-known group A of behaviours (both harmless and harmful combined) which we want to observe. Even if we want to label "unexplained" subsequences in a poorly characterised observation sequence (For instance, they can include instances of actions that have never been witnessed or expected, therefore they do not fall under category A).We correctly define the likelihood that a series of explanations is unexplainable (completely or partially) in terms of A. To find the top k entirely and partially unexplained sequences with respect to A, we advance effective methods. We can rapidly accelerate the search for completely or partially unexplained sequences using the strength of these algorithms' proposals. We construct tests utilising data from cyber-security and real-world video sets, demonstrating how our technique mechanism shines in training in terms of the correlation between running time and accuracy.

DOI Number: 10.48047/nq.2021.19.5.NQ21076 **NeuroQuantology 2021; 19(5): 303-307**

1. INTRODUCTION

Data mining is a young discipline that combines the realms of statistics, artificial intelligence, and databases. Numerous businesses have been able to gather vast volumes of data because to the information era. Yet, the value of this data is negligible if "meaningful information" or "knowledge" cannot be derived from it. This requirement is attempted to be satisfied via data mining, also known as knowledge discovery. Unlike traditional statistical methods, data mining approaches hunt for relevant information without requiring previous assumptions. It has developed as a field, introducing new ideas and techniques including association rule learning. It has also used renowned machine-learning techniques, such inductive rule learning, in situations involving very big databases. Data mining techniques are used increasingly in both corporate and academic settings.

A significant problem that emerges with any en masse collecting of data is confidentiality. Even while privacy is mandated by law in some cases (as with medical databases), business interests could also be at play. However, there are some situations where sharing data might be advantageous for both parties. Massive datasets are now mostly used for research, whether it be scientific, commercial, or market-oriented. Therefore, sectors like the medical field as well as competing businesses with similar objectives stand to gain greatly from data exchange for research. Despite the seeming advantage, confidentiality issues typically render this impracticable.

Finding unexpected behaviors is a key difficulty in a variety of applications, including cyber security, video surveillance, defect identification in safety-critical equipment, and fraud detection. For instance, video surveillance is used to continually check the luggage sections at airports for shady activity. Police frequently use video surveillance to keep an eye on parking lots and roadways in crime-ridden areas. In Israel, a centralized organisation keeps an eye out for any questionable activity on the roadways.Yet, all of these apps look for well-known behaviors—behaviors that have been previously classified as either risky or harmless. For example, in the highway application, security agents may watch for both "regular" conduct and "suspect" behavior, such as adhering to a certain speed limit unless traffic is moving slowly (e.g., A parcel was taken out and left on the side of the road

after the automobile had stopped close to a bridge before continuing on its trip). Cyber security intrusion detection can monitor network traffic for questionable activity and issue security warnings. Alerts are combined into multi-step assault situations using alert correlation methods. Even though, each method is dependent on models that store knowledge of either good or bad behavior that is known in advance.

Unusual occurrences like "zero days" assaults are too difficult for them to handle. In reality, none of these approaches can measure how well a set of incidents seen throughout an observation stream can be explained by the models that are currently available. In real life, our plan would function. We begin with a collection of "excellent" and "poor" activity model sets, A. Bad behaviors are those that are recognized to be wrong, whereas good behaviors are those that are thought to be suitable (for example, some approved actions at an airport secure luggage zone) (such as when a baggage worker opens a suitcase, removes the contents, and places them in a separate bag).There are existing methods for finding instances of certain behaviors in time-stamped observation data (such as a video, a series of online transactions, etc.), each of which has a corresponding probability.

2. LITERATURE SURVEY

Because to its capacity for learning from data and handling time-varying signals, hidden Markov models (HMMs) are often utilised in speech and vision. They are a poor model for voice recognition, though, given the majority of relevant signals do not meet the rigorous Markov requirement. In order to overcome this, coupled HMMs (CHMMs), which couple HMMs with temporal, asymmetric conditional probabilities, have been introduced. This approach offers advantages over Cartesian product HMMs, including the explicit representation of interactions between HMMs and the ability to record system behaviours in data. Unfortunately, it is inefficient in terms of time consumption and accuracy.The key benefit of this technique is that, unlike Cartesian product HMMs, it directly depicts interactions between HMMs, allowing it to capture system activity in the data. Regarding time usage, it is inefficient and inaccurate [1].

The requirement for an aggregate and correlation component (ACC) that may use warnings produced by intrusion-detection investigations is explored in this study. A single data model and a set of rules to process the alerts make up the two primary components of the ACC. The notion of circumstances enables the operator to aggregate related warnings and present a more condensed picture of the security risks to be handled, and the AC algorithm may detect duplication and repercussions. The ACC has a UI that is comfortable for the operator because it is founded on the Tivoli Enterprise Console and interfaces with a current event management system. They did not replace the more general AC algorithm rules with more particular ones. By doing alert normalisation as a front-end to the TEC, which is not addressed, performance can be improved [2].

This research proposes a methodology for identifying human activity using particle filter tracking and DBN-based identification in various contexts. It tracks scene objects and extracts spatio-temporal characteristics, such as the relative distances of items and agents, using a particle filter based on colour and shape. DBNs are trained using dynamic Bayesian networks before being applied to recognition. The intricacy of the items being tracked and the sequence of observed activities varied in each scenario, hence it was purposefully decided to conduct the experiment on a variety of situations with varying degrees of difficulty.The system is not reliable since they did not investigate and incorporate into to tell when full occlusion to dynamically collect data association [3].

This article examined the association between intrusion alarms in real time. Here, restrictions in using the nested loop-based correlation method are recognised, and an unique QG strategy is suggested to overcome these restrictions. The method can correlate warnings which are arbitrarily far apart and has a linear time complexity and quadratic memory demand. Here, expand the QG strategy to a unified methodology for the correlation, hypotheses, and prediction of alerts. Moreover, they expand technique for creating a condensed form of the finding graphs. According to empirical findings, correlation engines can process warnings more quickly than IDS can report them, making this technique viable way for administrators to keep track of the development of intrusions.They didn't test the methods on live networks with actual traffic. The nested-loop approach's drawback has not been solved, particularly with regard to correlating intense alarms on high-speed networks [4].

This research offered a unique method for body-worn sensors to identify an anomalous user activity. It employs a two-phase abnormality detection approach where a one-class SVM is constructed on normal activities and questionable trails are forwarded to a group of abnormal activity models that have been customised by KNLR. The method can better balance the detection rate and false alarm rates, but there is a chance that it will produce a lot of abnormal models if an aberrant action suddenly starts to become the norm. Moreover, the suggested strategy does not include integrating intelligent planning for courses of action when an abnormal scenario is detected

utilising statistical approaches or identifying aberrant behaviours from continuous user traces [5].

3. PROPOSED SYSTEM

To identify the odd behaviours in the present, several observations are needed. Based on a variety of characteristics, the intrusion detection system monitors network traffic. Signature-based and profile-based intrusion detection techniques are ineffective ways to discover a better solution. The packet headers and payload content are specified using the aforementioned two ways. Finding anomalies takes a long time because of the volume of observations. Intrusion detection may watch network traffic for questionable activity & issue security warnings in the context of cyber security. Methods for alert correlation combine alerts into multi-step assault situations. Both methods, though, depend on models that store beforehand information about good or bad conduct. They are unable to respond to unforeseen circumstances like "zero day" assaults.

We take into account the collection of known activities from typical activists in the proposed system. The basic objective of correlation is to offer a more comprehensive picture of the real assault, and correlation techniques are employed to identify causal relationships. It uses a domain-independent method to find the surprising sequences. Our goal in this study is to develop an unexplained sequence detector, which will allow us to pinpoint specific observation data that established models cannot "explain" with sufficient confidence. In other conflicts, the attainable activity models in A do not adequately capture what is happening in unexplained orders. Such sequences might then be further investigated, for example to create original activity models, after being discovered. The following is a list of the several benefits of the suggested strategy:

- The abnormality may be quickly identified by combining different data sets.
- Scalability is enhanced and accuracy is assessed.

Fig 1: System Architecture

The following section discusses the many stages involved in implementing the suggested system:

1. Dataset preprocessing

Data pre-processing is discussed in this module as a crucial phase in the data mining process. It entails putting unprocessed data into an intelligible format. The dataset must first be browsed and loaded into a database before being extracted.

2. Correlation Calculation

The goal of correlation in this module is to identify causal connections between alerts in order to reconstruct attacks from discrete alerts. Correlation's primary purpose is to provide a higher-level perspective of the actual assaults. Correlation techniques are ineffective in dealing with occurrences that are not explained by the underlying models because they rely on models that encode a priori knowledge of either typical or malicious behaviour.

3. Activities Classification

The gathered sets of known and unknown actions are categorised in this module. The activities here refer to a person visiting an ATM, inserting a card, and the number of tries that are being tracked. We need to categorise the process in light of the potential for temporal variation. We've categorised two actions here:

a. Known Activities

The known activities that a person engages in on a regular basis.

b. Unknown Activities

The unknown activities that are odd behaviours from a person.

4. Performance Evaluation

This module's evaluation compares and illustrates the current and suggested approaches using a graphical representation. Also, it displays a dataset's classification of normal and pathological behaviours.

4. RESULTS

The challenge of identifying inexplicable sequences in a domain-independent manner is examined in this work. It begins with a known set of behaviours (which are both harmless and harmful) and classifies them as "unexplained" subsequences in a poorly defined observation sequence. , the child of a tee-mark. Her eld. Finding an unexplained sequence detector, or identifying subsequences of the observation data

that established models cannot "explain" with a given degree of confidence, is the objective. Such subsequences can be further examined after being discovered, for instance, to derive novel activity models from them.

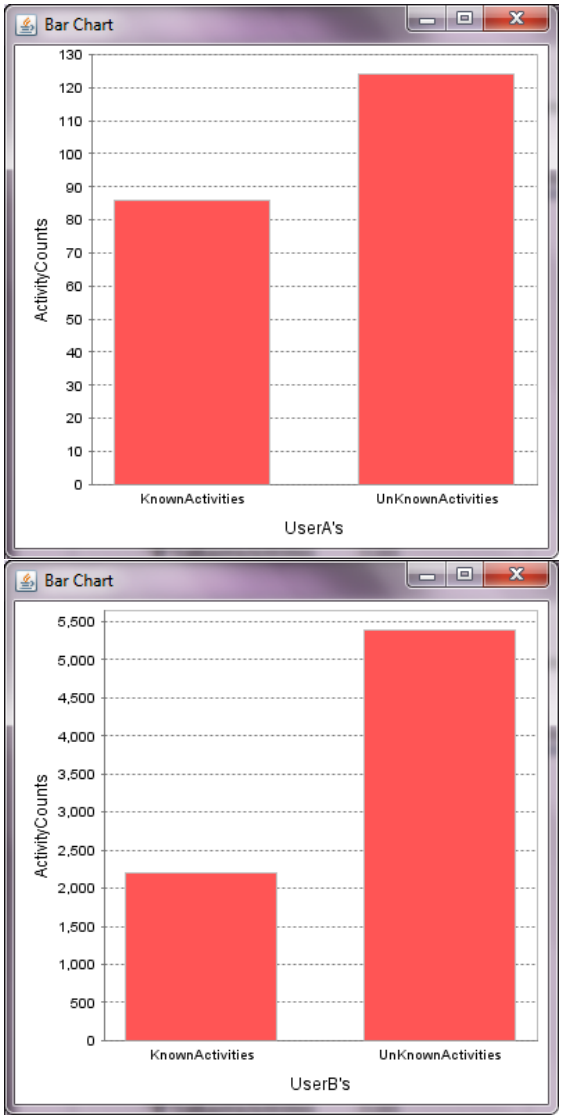


Fig 2: Performance Analysis

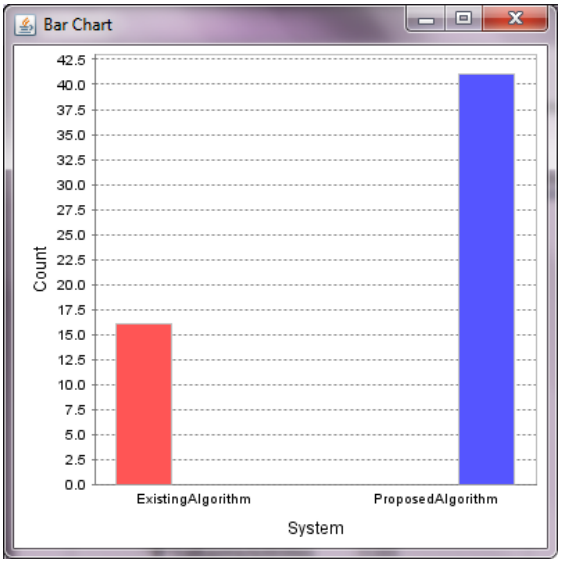


Fig 3: Comparative Analysis

5. CONCLUSION

The definitional consequences of v that cannot be "sufficiently" described by the actions in A are the subject of this essay. By analysing significant wholly and partially unexplained sequences, we correctly define what it means for a sequence to be unexplained. We identify fascinating traits that may be used to make the search for inexplicable sequences exceedingly efficient through intelligent pruning, and we propose a possible world's outline. The start of inexplicable sequence detection in a domain-independent manner is described in this study. Our objective in this study is to develop an unexplained sequence detector, or to recognise subsequences of the observation data that established models cannot "explain" with a given degree of confidence.In other words, the activity models in A that are now available do not adequately describe what occurs in unexplained sequences. These subsequences can then be further examined, for instance, to derive novel activity models from them, after they have been discovered.

6. FUTURE ENHANCEMENT

We enhance this technique by employing association rule mining in our subsequent work to categorise people's activities. If/then clauses known as association rules can be found in transactional databases, relational databases, or other information repositories to assist identify connections between seemingly unrelated data.

REFERENCE

[1] M. Albanese, V. Moscato, A. Picariello, V.S. Subrahmanian, and O. Udrea, “Detecting Stochastically Scheduled Activities in Video,” Proc. 20th Int’l Joint C onf. Artificial Intelligence, pp. 1802-1807, 2007.

[2] S. Hongeng and R. Nevatia, "Multi-Agent Event Recognition," Proc. Int'l Conf. Computer Vision (ICCV), pp. 84-93, 2001.
[3] N. Vaswani, A.K.R. Chowdhury, and R. Chellappa, "'Shape Activity': A Continuous-State HMM for Moving/Deforming Shapes with Application to Abnormal Activity Detection," IEEE Trans. Image Processing, vol. 14, no. 10, pp. 1603-1616, Oct. 2005.
[4] M. Brand, N. Oliver, and A. Pentland, "Coupled Hidden Markov Models for Complex Action Recognition," Proc. IEEECS Conf. Computer Vision and Pattern Recognition (CVPR), pp. 994-999, 1997.
[5] N. Oliver, E. Horvitz, and A. Garg, "Layered Representations for Human Activity Recognition," Proc. IEEE F our th Int'l Conf. Multimodal Interfaces (ICMI), pp. 3-8, 2002.
[6] R. Hamid, Y. Huang, and I. Essa, "Argmode - Activity Recognition Using Graphical Models," Proc . Computer Vision and Pat tern Recognition Workshop (CVPRW), pp. 38-43, 2003.
[7] N.P. Cuntoor, B. Yegnanarayana, and R. Chellappa, "Activity Modeling Using Event Probability Sequences," IEEE Trans. Image Processing, vol. 17, no. 4, pp. 594-607, Apr. 2008.
[8] D. Zhang, D. Gatica-Perez, S. Bengio, and I. McCowan, "Semi-Supervised Adapted HMMs for Unusual Event Detection," Proc. IEEECS Conf. Computer Vision and Pattern Recognition (CVPR), pp. 611-618, 2005.
[9] M.T. Chan, A. Hoogs, J. Schmiederer, and M. Petersen, "Detecting Rare Events in Video Using Semantic Primitives with HMM," Proc. 17th Int'l Conf. Pattern Recognition (ICPR), pp. 150-154, 2004.
[10] T. Xiang and S. Gong, "Video Behavior Profiling for Anomaly Detection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30, no. 5, pp. 893-908, May 2008.
[11] J. Kim and K. Grauman, "Observe Locally, Infer Globally: A Space-Time MRF for Detecting Abnormal Activities with Incremental Updates," Proc. Conf. Computer Vision and Pattern Recognition (CVPR), 2009.
[12] J. Yin, Q. Yang, and J.J. Pan, "Sensor-Based Abnormal Human-Activity Detection," IEEE Trans. Knowledge and Data Eng., vol. 20, no. 8, pp. 1082-1090, June 2008.
[13] D.H. Hu, X.-X. Zhang, J. Yin, V.W. Zheng, and Q. Yang, "Abnormal Activity Recognition Based on Hdp-HMM Models," Proc. Int'l Joint Conf. Artificial Intelligence, pp. 1715-1720, 2009.
[14] X.-X. Zhang, H. Liu, Y. Gao, and D.H. Hu, "Detecting Abnormal Events via Hierarchical Dirichlet Processes," Proc. 13th Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining (PAKDD), pp. 278-289, 2009.
[15] D. Mahajan, N. Kwatra, S. Jain, P. Kalra, and S. Banerjee, "A Framework for Activity Recognition and Detection of Unusual Activities," Proc. Indian Conf. Computer Vision, Graphics & Image Processing (ICVGIP), 2004.
[16] F. Jiang, Y. Wu, and A.K. Katsaggelos, "Detecting Contextual Anomalies of Crowd Motion in Surveillance Video," Proc. IEEE 16th Int'l Conf. Image Processing (ICIP), pp. 1117-1120, 2009.