Hypergraph based Unsupervised Contextual Pattern Learning and Anomaly Detection for Global Terrorism Data

Akanksha Toshniwal1, Kavi Mahesh2 and Jayashree R3

1PhD Scholar, PES University, Bangalore, India.
2Director, IIIT, Dharwad, India.
3Professor, PES University, Bangalore, India.

Email: akanshatoshniwaleca@gmail.com

Abstract. In a dataset, an event which deviates from the rest of the dataset is a rare event. This rare event can be intrusion or any suspicious activity in the system and is called an anomaly. These anomalies are important to detect because this may be any terrorist attack, outbreak of the disease, malfunctioning or fraud in the system. Anomalies are the deviation from the normal patterns in the dataset. It is important to learn the normal patterns in order to identify the deviation. Labelled data in real life anomaly detection is not available due to rarity of anomalies. It is challenging to identify anomalous combinations and combinatorial patterns of feature instances using conventional machine learning algorithms. We introduce Hypergraph based Unsupervised Contextual Pattern Learning and Anomaly Detection (HUCPLAD) technique for unlabeled datasets and implemented on Global Terrorism Data (GTD). HUCPLAD gets rid of the curse of dimensionality, maintains hierarchy, learns the contextual pattern, detects contextual anomalies and measures the behavior of co-occurring events.

1. Introduction

There are three types of anomalies. Point Anomaly: When an individual data sample is anomalous or unusual with respect to the rest of the pattern. e.g. sudden drop and increase in sales or temperature, Collective Anomaly: When a collection of some related activities in a dataset are anomalous with respect to the entire data set. e.g. Continuous peak or no signals in an electrocardiogram machine. Contextual Anomaly: When a particular event is anomalous with one context but normal with another context or when occurs individually, e.g. attributes: temperature (Centigrade), location (city) and season (summer,winter) with instances (12, Jaipur, Winter), (36, Jaipur, Summer), (36, Jaipur, Winter). The temperature, location and season values are not anomalous individually. Temperature of 36 Degree C is normal in Summers in Jaipur, but it is abnormal for 36 Degree C in Winters in Jaipur. Same data instance has anomalous and normal behaviour with different contexts. Here (location, season) are contextual attributes and (temperature) is behavioural attribute. Pattern learning for combination of these co-occurring instances is necessary in order to detect anomalous combinations.

Here we are interested to detect the normal and anomalous patterns with respect to some context. There are two types of attributes in contextual anomaly detection 1. Contextual attribute: defines the context of the event. 2. Behavioral attributes: defines the behavior of the context.

Contextual patterns and anomalies in the dataset, may remain unrecognized if used traditional machine learning or graph technique [1] [7].

This paper proposes pattern learning and anomaly detection by utilizing properties of hypergraph and statistics using a similar approach as anomaly detection and neighborhood formation in bipartite graphs in [7]. Limitation to [7] is that the relationships between the graph nodes are defined pairwise, but in
hypergraph these relationships can be defined in groups of vertices of any size.

Hypergraph represented as $H = (V, HE)$ is the generalization of the graphs where $V$ is a set of vertices/nodes and $HE$ hyperedges are the generalization of edges, subsets of vertices $V$. Let every instance $V = \{x_1, x_2, x_3, \ldots, x_p\}$ be a finite set where $x_i \in V$ where $i = 1, 2, \ldots, p$ and $p$ is the total number of nodes and let $HE = \{he_1, he_2, he_3, \ldots, he_n\}$ be a family of subsets of $V$. A simple hypergraph is represented in Figure 1 and its bipartite graph is shown in Figure 2.

![Hypergraph](image1.png)

**Figure 1.** Hypergraph representation for Hypergraph: $H=(V, HE)$, where Hyperedge set $HE = \{he_1, he_2, he_3, he_4, he_5\}$ and Vertex Set instance $V = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$.

![Bipartite Graph](image2.png)

**Figure 2.** Bipartite Graph $B(H)$ representation of $H(V, HE)$, $V_1=\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$, $V_2=\{h_1, he_2, he_3, he_4, he_5\}$, $E$ is edge set connecting $V_1$ and $V_2$. 
2. Related Work
Pattern learning and anomaly detection goes hand in hand. Quasi real time anomaly detection and pattern learning on time series energy management of buildings is proposed in [2] where enhanced Symbolic Aggregate approximation (SAX) process is used to reduce time series without losing key information. Deep learning graph-based anomaly detection [14] using sum product networks (SPN) is proposed in [13] on directed acyclic weighed graph. Various unsupervised anomaly detection and pattern learning techniques like autoencoder, isolation forest, nearest neighbor, K Means, scaled convex hull, OCSVM are compared in [15] on KDD [16] and ISCX [17] datasets for cyber-attacks where 1-Nearest Neighbor was capable of detecting most of the anomalies and fastest among other unsupervised techniques.

The hypergraphs help to find optimal solutions for many new optimization problems [3]. Anomaly detection is one of the applications which we utilize the hypergraph structure. There are traditional applications for anomaly detection [1] and there are graph-based anomaly detection [5] but the limitation with traditional anomaly detection is that they are unable to identify the context of the anomaly, and graph based methods fails to show the connection among multiple nodes because an edge can only connects two nodes [6][7]. Graph based neighborhood formation or search for similar nodes and anomaly detection is proposed in [7], where normality score is computed for identifying anomalies in a graph by modeling into bipartite graph and the relevance score represents neighbors of nodes (nearby neighbors). Both scores are calculated on graph nodes.

Anomaly detection using scan statistics is performed on directed graph in [8] and on undirected hypergraph in [9], implemented on an Enron time series dataset. The idea is to compute the locality statistic at location as vertex for different order neighborhoods and event with scan statistics greater than five standard deviation above mean the event is called as anomaly. High dimensional anomaly detection is proposed by [10] and anomaly detection in very large network is proposed in [11] by representing on hypergraph as binary strings (0s and 1s) of size equal to count of unique vertices in hypergraph where data is identically independently distributed. Variational approximation to probability mass function is proposed to get variational expectation maximization which estimates the parameter for distribution. The proposed technique is similar to the graph-based technique in [7] where neighborhood formation is computed on bipartite graph.

3. Problem Definition
We propose HUCPLAD for discrete dataset. Hypergraph $H = (V, HE)$ with ‘p’ nodes and ‘n’ observations. Each hyperedge is the subset of the vertex set powerset. The aim of this technique is to identify the behavior of the co-occurrences of hyperedges in an event. The following definitions are defined:

**Definition 1:** Given a hyperedge of cardinality $|HE_i| \leq 1,2...n$ the vertex set associated with $HE$ is set $i = \{1,2,3...n\}$. Then the primary values are computed as:

$$I(w,j) = \{ HE_{id} \in w \mid I(i,j) = 1; x_i \in HE_j; w \in x_i, where i = 1,2,..p,... . . (1)$$

### Table 1. Incidence matrix I(H) for given Hypergraph H.

| id | x1 | x2 | x3 | x4 | x5 | x6 | x7 | x8 | x9 |
|----|----|----|----|----|----|----|----|----|----|
| he1 | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  |
| he2 | 0  | 0  | 1  | 1  | 1  | 0  | 0  | 0  | 0  |
| he3 | 1  | 0  | 0  | 0  | 1  | 1  | 0  | 0  | 0  |
| he4 | 0  | 1  | 1  | 0  | 1  | 1  | 1  | 0  | 0  |
| he5 | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 1  |
Where, Incidence Matrix, \( I(i,j) = 1 \) if \( x_i \in HE_j \), 0 if \( x_i \notin HE_j \). (2)

Which means, the primary value is defined as hyperedge ids for which incidence matrix value is true for all \('w'\) sets of vertices.

**DEFINITION 2:** Given hyperedge set of cardinalities \( |HE_1^{i}, |HE_2^{i}, |HE_3^{i}, \ldots |HE_n^{i} \mid \geq 1, 2, \ldots, n \) and the context hyperedge \( |HE_{i}^{\text{context}}| \) of any cardinality. The hyperedge behavioural score in context with \( |HE_{i}^{\text{context}}| \) is calculated as:

\[
\text{Behavior score} = \frac{\text{Numerator}}{\sigma} \quad \text{…………………}(5)
\]

Where, Numerator\(=|I^{c.b}(w,j)| - \mu \quad \text{…………………}(6)\)

\[
\sigma = \sqrt{\frac{\sum (|I^{c.b}(w,j)| - \mu)^2}{n}} \quad \text{……………………………….}(7)
\]

\[
I^{c.b}(w,j) = \{1|\text{x}_i \in HE^{\text{contextual behavior}}, w \text{EX}_i\} \quad \text{………………………………}(8)
\]

\[
\mu = \frac{\sum_{\text{behavioral he cu contextul he}} I^{c.b}(w,j)}{n} \quad \text{…………………}(9)
\]

\(w \in x_i\), where \(i = 1, 2, \ldots, p\). Behavior score is the normalized standard score computed on every contextual hyperedge with each behavioral hyperedge in hypergraph.

**DEFINITION 3:** Given a hyperedge of cardinality \( |HE_1| = 1, 2, \ldots, n \), the hierarchy level of hyperedge is defined as:

\[
L(HE) = \{ |HE_1|^{\text{Contextual}} + |HE_1|^{\text{Behavioral}} \mid \text{for distinct } x_i \} \quad \text{………………………}(10)
\]

Hierarchy level is sum of cardinality of contextual and behavioural for distinct vertex set in hyperedge.

Given a context hyperedge and behavioral hyperedge, the frequent patterns are learnt is called as pattern learning. Given a context hyperedge and behavior hyperedge, whether the pattern is anomalous is called contextual anomaly detection. Objective of this paper is contextual pattern learning and contextual anomaly detection.

### 4. Proposed Method

This section discusses about the proposed algorithm, assumptions, and delivers following statements:

- It is an unupervised approach to detect patterns on unlabeled dataset.
- It provides dimensionality reduction without losing the data.
- Does not require transforming discrete data instances to continuous.
- It detects contextual patterns and contextual anomalies.
- Maintains the hierarchy for each context.
- It deals with missing fields seamlessly.

The assumptions are:

- The dataset is well structured.
- The dataset is unlabeled.
- Anomalies are rare.
Anomalies are different than the norm.
The data instances in dataset are mostly discrete.

Algorithm HUCPLAD generates the hypergraphs using association rule mining with hyperparameters support confidence and frequency. Each hyperedge has a hierarchy with some cardinality. Behavioral score is calculated for each cell in adjacency matrix $M_A$ by computing the incidence matrix (see table 1) on bipartite graph (see figure 2) of the given hypergraph $H = (V, HE)$. Each cell in output matrix $M_A^{behavior score}$ is the behavioral score for contextual hyperedge (row) and behavioral hyperedge (column).

Algorithm 1. HUCPLAD Algorithm.

| Algorithm 1: HUCPLAD | $H_{DB}$: Dataset |
|----------------------|--------------------|
| **Input:** DB, t_confidence, t_frequency | HE: Hyperedge |
| **Output:** $M_A^{behavior score}$ | Context: Contextual attributes |
| | Behavior: Behavioral attribute |
| | t_confidence: confidence_threshold |
| | t_freq: threshold_frequency |

1. Association Mining ($H_{DB}$, t_confidence, t_freq)
2. Build the hyperedges $HE$ and construct the hierarchy for each $HE$.
3. Construct Adjacency Matrix $M_A$ of size $j \times j$ on $HE = (HE_{combinational})$.
4. Behavior Score $M_A$

Algorithm 2. Behavior Score Algorithm

| Algorithm 2: Behavior_Score | $M_A^{behavior score}$ |
|-----------------------------|-------------------------|
| **Input:** $M_A$ | $M_A^{behavior score}$ |
| **Output:** $M_A^{behavior score}$ | score with behavior score |

1. For context/row in $M_A$:
   a. For behavior/column in $M_A$:
      i. Compute $I_{cb}(w, j)$ for every cell.
      ii. Compute for each context.
      iii. Compute sigma $\sigma$ for each row.
      iv. Compute behavior score for each cell.
   b. End For
2. End For
3. Return $M_A^{behavior score}$ score with behavior score.

5. Experiment and Results

Experiment’s specification: Operating System Linux Ubuntu 16.04; RAM 16GB; processor i7 CPU@2.6GHz; Language: Python 3, Libraries: pandas, numpy, apyori, pyfpgrowth; Anaconda 3, multiprocessing in Linux environment; Run Time: 17 Hours to 24 Hours. Hyperedge selection is done by implementing association rule mining which limits the maximum cardinality.

The experiment is performed on Global Terrorism dataset [12], there are two contexts global context (static) and local context (dynamic). The experiment considered 13 features [State, year, city, specificity, attack_type, targ_type, targ_subtype, corp, target, gname, weap_type, weap_subtype, number_killed] for country India and state Maharashtra for duration 2012 to 2017. For global context: country: India, state: Maharashtra, duration: 2012 to 2017 and local context= number of people killed is exactly 1 which could
be target attack on particular person.

Each event has a frequency and a behavior score. In figure 3 it is observed that the frequency and behavior score are correlated in such a way that higher frequency denotes higher score and lower frequency denotes lower score. Very lower score is the hint that the event is most likely to be anomalous. Each figure 3a, 3b, 3c, 3d, 3e is score vs. frequency (count) plot, where x-axis denotes the behavioral hyperedges and its cardinality. The gradual decrement in score can be noticed as the frequency (count) decreases. Distribution of the behavior score is shown in figure 4 in the box plot. Some of the following results for global context: country: India, state: Maharashtra, duration: 2012 to 2017 Local Context: nkill=1.0 are obtained from the plot:

- City= {Sirpur, Koparshi, Menda, Bijepar}, WeaponSubType ={Granade, Arsonfire} are anomalous.
- Gname = {Maoists} AND weapontype = {Incediary}, AttackType = {Facility Infrastructure Attack} AND weapontype = {Incediary} are anomalous.
- Gname = {Maoists} AND target type= {Police} AND target sub type= {Police building headquarter station school} is an anomaly.
- Iyear= {2012} AND nkill= {0} AND specificity= {1} AND Weapon type= {Explosive} is an anomaly.

Figure 3 (a). Score and Frequency plot for behavioral hyperedge Degree = 1.
Figure 3 (b). Score and Frequency plot for behavioral hyperedge Degree =2.

Figure 3 (c). Score and Frequency plot for behavioral hyperedge Degree =3.
Figure 3(d). Score and Frequency plot for behavioral hyperedge Degree = 4.

Figure 3(e). Score and Frequency plot for behavioral hyperedge Degree = 5.
Figure 3 Score and Frequency (count) plot for global context: country: India, state: Maharashtra, duration: 2012 to 2017 Local Context: nkill=1.0 where score \{R+,R-\} and count \{Z^{a0}\}

Figure 4. Box plot for Score when context is nkill=1.0, this box plot represents the distribution for all hyperedges.

6. Discussion
HUCPLAD is a score-based technique which detects the normal or anomalous contextual patterns in largely discrete dataset. This technique requires to find all the combinations in the dataset, these combinations are called as hyperedges in the hypergraph, where each hyperedge has some cardinality. But working with all the combination is very costly with respect to time and memory, so we propose using associate rule mining which ensures no data loss. Behavior score is calculated for each cell in the adjacency matrix, the score and frequency are positively correlated. Lower frequency indicates low behavior score and vice versa. The hyperedges with very low scores (less than 0) are more likely to be anomalous. However, the future work is to find the way to define a threshold which clearly specifies a pattern’s class.

7. Conclusion
This paper proposed the hypergraph based unsupervised contextual pattern learning and anomaly detection (HUCPLAD) on largely discrete dataset. The experiment is performed on Global Terrorism dataset where two different contexts are defined “global context” which is static and “local context” which is dynamic, there are 1194 different dynamic local contexts associated with single global context and for each local context 1194 different behaviors are associated. For each dynamic local context and behavior the proposed method creates 1194 x 1194 matrix, where each cell of this matrix defines behavior score. The behavior score represents whether the event is anomalous or non-anomalous. The anomalous events are those hyperedges combinations that has never occurred in the dataset. The behavior score is important because when the maximum frequency of the behavioral hyperedges is unknown, the score is useful to identify whether the pattern is normal or anomalous.

References
[1] Chandola, Varun, Arindam B, and Vipin K 2009 Anomaly detection: A survey ACM computing surveys (CSUR) 41.3 1-58.
[2] Capozzoli, Alfonso, et al. 2018 Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings Energy 157 336-352.
[3] Voloshin, Vitaly I 2009 Basic Hypergraph Concepts Introduction to graph and hypergraph theory Nova Science Publ. 135-154.
[4] Meira, Jorge, et al. 2020 Performance evaluation of unsupervised techniques in cyber-attack anomaly detection Journal of Ambient Intelligence and Humanized Computing 11.11 4477-4489.
[5] Akoglu, Leman, Hanghang T, and Danai K 2015 Graph based anomaly detection and description: a survey Data mining and knowledge discovery 29.3 626-688.
[6] Toshniwal A, Mahesh L and J R 2020 Overview of Anomaly Detection techniques in Machine Learning 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India 808-815.

[7] Sun, Jimeng, et al. 2005 Neighborhood formation and anomaly detection in bipartite graphs Fifth IEEE International Conference on Data Mining (ICDM05) IEEE.

[8] Priebe, Carey E, et al. 2005 Scan statistics on enron graphs Computational & Mathematical Organization Theory 11.3 229-247.

[9] Park, Youngser, et al. 2009 Anomaly detection using scan statistics on time series hypergraphs Link Analysis, Counterterrorism and Security (LACTS) Conference. SIAM.

[10] Silva, Jorge, and Rebecca W 2008 Hypergraph-based anomaly detection of high-dimensional co-occurrences IEEE transactions on pattern analysis and machine intelligence 31.3 563-569.

[11] Silva, Jorge, and Rebecca W 2008 Hypergraph-based anomaly detection in very large networks.

[12] GTD Access https://www.start.umd.edu/gtd/access/, 2020 (accessed 10 August 2020).

[13] Peharz, Robert, et al 2018 Probabilistic deep learning using random sum-product networks arXiv preprint arXiv:1806.01910.

[14] Chalapathy, Raghavendra, and Sanjay C 2019 Deep learning for anomaly detection: A survey arXiv preprint arXiv:1901.03407 (2019).

[15] Meira, Jorge, et al. 2020 Performance evaluation of unsupervised techniques in cyber-attack anomaly detection Journal of Ambient Intelligence and Humanized Computing 11.11 4477-4489.

[16] Tavallaee, Mahbod, et al. 2009 A detailed analysis of the KDD CUP 99 data set 2009 IEEE symposium on computational intelligence for security and defense applications. IEEE.

[17] Shiravi, A., Shiravi, H., Tavallaee, M., & Ghorbani, A. A. 2012. Toward developing a systematic approach to generate benchmark datasets for intrusion detection. computers & security, 31(3), 357-374.