M.S. Dissertation in Engineering

Machine Learning for Anomalous Human Insurgency Detection

- Deep Neural Network for Anomaly Detection-

비이상적 무력 폭등 감지를 위한 기계 학습
: Deep Neural Network 을 활용한 비이상적 징후 감지를 중점으로

August 2019

Graduate School of Seoul National University

Technology Management, Economics, and Policy Program

Sul Lam Jeoung

Machine Learning for Anomalous Human Insurgency Detection

지도교수 황준석

이 논문을 공학석사학위 논문으로 제출함

2019 년 8 월

서울대학교 대학원

협동과정 기술경영경제정책 전공

정술람

정술람의 공학석사학위 논문을 인준함

2019 년 8 월

위 원 장 **Jörn Altmann** (인)

부위원장 황 준 석 (인)

위 원 윤 현 영 (인)

Abstract

Machine Learning for Anomalous Human Insurgency Detection

Sul Lam Jeoung
M.S. Technology Management, Economics and Policy
The Graduate School
Seoul National University

Human insurgency is one of the prevalent, incessant, and threatening events happening worldwide. Among many topics of developmental studies, one of the seminal research focuses is to understand and model armed conflicts, which have been suspected to be linked to the capacity of a country in various ways, such as food security, child nutrition, economic welfare, and even environmental issues. Mapping human insurgencies is, therefore, imperative.

To cope with the atrocities, there have been previous attempts to uncover the latent patterns of human insurgent incidents. The salient behavior of these insurgencies follows the 'power-law' distribution, which exhibits a heavy-tail. This feature implies that events far from the norm are nontrivial when compared with the normal distribution, where essentially no weight is far from the mean. This pattern indicates that the insurgencies are the few incidents happening with relentless severity, while the majority of the events occur with mere severity. To fully exploit the latent behavior of human insurgencies, this research focuses on the “anomalies” — the events that have a great number of fatalities but little probability of occurrence, lying on the heavy tail—.

To detect such anomalies, a novel approach, variational autoencoder, is used. The seminal essence of this model lies in processing high-volume data and capturing their non-linearity, which makes data-driven detection possible. The results show that the trained model successfully detects anomalies when given test data, showing no false negatives (Type III error) or false positives (Type I error). This predictive model, if well deployed, can provide humanitarian aid agencies and governments the ability to efficiently allocate resources, reducing wastes and mitigating the level of conflict through targeted preventive policies.

Keywords: VAE, Machine Learning for development (ML4D), Anomaly Detection, Human insurgency, Armed conflict.

Student Number: 2017-28957

Contents

List of Tables

List of Figures

Chapter 1. Introduction

The violent events have been incessant and prevalent. There have been ceaseless massive bloodsheds in perilous areas. The year 2017 was recorded as one of the most violent years since the end of the Cold War. Afghanistan, for example, witnessed its most violent year and Iraq second most violent year in 2017, during the post-Cold War period (Dypuy, K et al.,2018). According to the report from United Nations, many regions of the world suffer from continuous untold horrors as a result of armed conflict or other forms of violence. Mapping crisis, therefore, alleviating the level of violence has been designated as one of the sustainable development goals by the United Nations.

The armed conflicts have been assumed for having linkages between interstate, and intrastate issues, such as food security, state capacity, child nutrition, economic welfare: poverty, and environmental issues. (Martin-Shields,C.P et al., 2019; Kim.H,2019; Odozi &Oyeler,2019;,Breckner,2019). To prevent and cope with the atrocities, crisis mapping has been in need than ever (Ziemke,2012) and among many topics of developmental studies (e.g. social, economic, environmental studies), a critical objective in institutional focus is to model violence (De-Arteaga, M et al, 2018). Predictive models could provide humanitarian aid agencies and governments the ability to efficiently allocated resources and reduce the future level of conflict.

There have been successive and abound attempts to develop predictive models. However, the novel approach is imperative. Firstly, the dynamics of these incidents have become more complex and disorderly. According to the report published by PRIO, "Trends in Armed conflict 1946-2017", the dynamics of armed conflict has changed. The world is less deadly compared to the Cold War era. Large scale warfare has declined compared to the 20th century, where fewer people are relentlessly killed. While cross-border, interstate conflicts have waned, however, non-state violence increased dramatically. The dynamics of conflicts has become a more disorderly and more complex issue. One of the possible reasons for this can be explained by small groups of so-motivated individuals able to gain access to destructive weapons, leading to the fragmentation and proliferation of armed groups, and an increase in non-state conflicts (Shubik 1997; Federal Bureau of Investigation 1999).

Secondly, the previous literature of predictive models and the latent behavior analysis are unfolded in a disparate manner. While considerable research uncovers the latent behavior of human insurgencies in a statistical way, the statistical analysis is often conducted per se. Some researches give simulations or toy models reflecting the latent behaviors, however, the predictive models which reflect this latent behavior empirically is scarce. The previous predictive models generally focus on the outbreak of the incidences. The latent behavior implies that there is a need to focus on the severity of incidences. The severity of such incidences exhibits a 'power-law', with the heavy tail. This implies that the heavy tail is nontrivial and non-negligible.

The research objective of this paper is to suggest a predictive model which incorporates the latent behavior of human insurgency based on spatiotemporal information. Based on this objective, the research question formulated is likewise: What is the implication of the latent behavior of human insurgencies? How can we reflect the latent behavior in the predictive model?

To develop this research question, this paper focuses on the anomalous event detection, incorporating the latent behavior of human insurgencies. To model the anomalous event, Variational Autoencoder(VAE) with deep neural network methodology has been used. It has been used for the following reasons. First, the theoretical foundations of VAE are firm and as one method of dimensionality reduction, it reduces dimensions in a probabilistically sound way. Secondly, with the efficacy of deep learning, VAE with stacked (deep) layers are known to enhance the representational ability of the data. Highly stacked hidden layers better extract abstract features to form a better reconstruction of the data. Thirdly, the characteristics of deep learning is often construed as 'Black Box', unable to analyze which covariate have led to such consequences (or prediction). The structure of VAE, however, allows to analyze the attributes that have led to such result. Lastly, the advantage of VAE over other methods (e.g. SVM classifier or Bayesian Classifier or PCA) is that it provides a probability measure as an anomaly score. This probability measure for VAE is more objective and principled since the performance of the model is not threshold sensitive. Unlike VAE, SVM classifier, for example, is sensitive to hyper-parameters where the specific threshold for judging anomalies is critical.

The predictive model was tested with the empirical data of Syria. This model can give a contribution to understanding and predicting the anomalous outbreaks, given spatiotemporal information. The structure of VAE also allows the covariate analysis of spatiotemporal data.

The remainder of the paper is organized as follows: Chapter 2 offers related literature and works. The explanation of the methodology and the data is presented in Chapter 3. Chapter 4 discusses the results of the analysis. Finally, Chapter 5 closes the paper with its implication, contribution, and limitation of this research

Chapter 2. Literature Review

2.1 Human Insurgency

The general term of human insurgency is defined as "an organized rebellion against widely recognized state authority when the perpetrators are not considered legitimate or appropriate under international rules of sovereign conduct of war" (Adelaja and George,2019; Morris,2005, Tomes and Robert R,2004). The tactics of this insurgency range from nonviolent activities such as propaganda, to violent activities such as bombings of civilian, government targets and critical infrastructure. Terrorism and subversion can be construed as one of the operational strategies of human insurgency, according to this definition. Recently, the term is more frequently used as a coined term to indicate specific events, such as Boko Haram Insurgency (BHI) of Nigeria, Maoist Insurgency of Nepal (Adelaja and George,2019, Jackson,2019; Nandwani,2019).

The definition of human insurgency is ambiguous in its nature since an internationally accepted term doesn't exist. There is a rigorous, discordant distinction between rebellion and belligerents. The stances of Third Geneva Convention and the United States Department of Defense (DOD), for example, differ by circumscribing it to only national states or loosely involving irregular forces / taking morality perspective into account or just focusing on the operational functions.

It is stated in The Third Geneva Conventions that "Members of other militias and members of other volunteer corps, including those of organized resistance movements, belonging to a Party to the conflict and operating in or outside their own territory, even if this territory is occupied, provided that such militias or volunteer corps, including such organized resistance movements..", limiting the actors of insurgency as nation-states and only loosely addressing the irregular forces. Meanwhile, the United States Department of Defense (DOD) defines the insurgents as "an organized movement aimed at the overthrow of a constituted government through the use of subversion and armed conflict." This definition is more focused on the operational actions, leaving the morality of the conflict aside. The subtle shades of difference aren't covered in this paper, which is beyond the scope of this paper.

The overall usage of the term, however, is used to indicate armed conflicts between actors, which entails armed conflicts and violent acts that engender deaths (Bohorquez, J.C., et al, 2009; Clauset,2012). The scrutiny of the purpose-whether political or religious- and strategies of the acts are marginalized. Bohorquez, 2009, for example, defines and quantifies human insurgencies, by only taking into account the violent conflicts which have a number of casualties.

2.1.1 Definition of Human Insurgency

In this paper, the term human insurgency incorporates the violent events of 'Battle' and 'Violence against civilians'. The categorization of the event 'Battle' and 'Violence against civilians' follows the definition from ACLED (Armed Conflict Location & Event Data Project).

The battle is defined as "a violent interaction between two politically organized armed groups at a particular time and location. It can occur between armed and organized state, non-state, and external groups, and in any combination therein". The term 'battle' may include different kinds of encounters. It even includes the state where a ceasefire is broken, however, it must involve violent events among at least two armed and organized actors. If it is a one-sided act, it is classified as violence against a civilian. The sub-events of battles are armed clashes, government against the territory, and non-state actor overtaking the territory. The analysis level was set aggregating the sub level distinctions.

Violence against civilian is defined as "deliberate violent acts perpetrated by an organized political group such as a rebel, militia or government force against unarmed non-combatants." Civilians are by definition unarmed, and unengaged in violence. Civilians are defined as unarmed non-combatants. The perpetrators of these acts incorporate its affiliates, militias, and other external forces. These acts not only involve attempts at inflicting harm but also forcibly disappearing (e.g. kidnapping and disappearances). The only incidents which the number of fatalities is at least one are counted as a valid event for analysis. The sub-events of violence against civilian are sexual violence, attack, abduction, and forced disappearance. The sub-events were aggregated for the level of analysis.

The distinction between Battle and Violence against civilians is that battle must involve at least two armed actors, while violence against civilians is a one-sided interaction. Both of them, however, must include violent interactions.

2.2 Predictive Models of Human Insurgency

The term 'prediction' is defined as to assign a probability distribution to events, given the model estimates from the observed data, while forecasting is to obtain probability distribution only limited to unrisen future events (Hegre et al.,2017). It has been one of the core task of peace research according to Singer (Singer,1973) and considerable research on prediction and forecast has been conducted (O'Brien, 2010; Brandt, Freeman, & Schrodt, 2011; Schrodt, Yonamine & Bagozzi, 2013) with a number of methodological approaches, for example, Game theory (Bueno de Mesquita, 2010), machine-learning tools(Schrodt, 1991) , automatic coding algorithm (Schrodt, Davis & Weddle, 1994).

The predictive modeling differs by variety of covariates it adopts. The general and basic approach to predictive modelling, has been focused to spatial and temporal components (Kupilik,2018). The spatial focus of modeling, also known as kernel intensity estimation, assumes that the hot spots where violent events happen frequently would also likely to continue in the near future(Eck,J et al, 2005;Kalyvas 2006 and Boone 2003;Höglund, K. et al,2016). It is frequently used in the forecasting context, even though it contains no temporal information. This method is reasonably accurate for the short term (Flaxman,2014).

The models of temporal focus assume that there is a seasonality that governs the occurrence of violent event. Univariate time-series forecasting models (Gorr,W.et al.,2003) including random walk and variety of exponential smoothing methods fall into this temporal analysis category. It considers no spatial component, which each is estimated separately for different locations. In general, predicting the highly frequent events is well performed, however, the events that occur rarely, the performance is not guaranteed. To overcome the limitations of basic spatial and temporal model, mixture of spatial and temporal models has been suggested. Such spatiotemporal is implemented in Gaussian Processes to model both the heat maps and temporal trends to forecasts. Space and time components are considered essential inputs to prediction, since conflicts and events exhibit strong temporal and spatial autocorrelation.

2.2.1 The limitation of previous literature

The predictive models specified above focuses on predicting the incidence of an event rather than the severity or the scale of such incidences. The spatiotemporal model suggested by Kupilik (Kupilik,M.,2018), for example, captures the spatial and temporal information and focuses on when and where the incidence would likely to take place. Likewise, the previous literature often disregarding the severity may give misguiding policy implication, such as insufficient preventive actions or myopic aftermath response. There is a need to take the severity of an event into account to give profound implications of a potential event. To deal with such issues, this paper covers the previous literature of latent behavior of human insurgency to incorporate the severity measure into the prediction

2.3 Latent Behavior of Human Insurgency

In the complexity science perspective, there have been attempts to uncover the latent patterns of human activities or social phenomenon. Lewis Fry Richardson, one of the founding fathers of modern complexity science explored how complexity emerge from the interaction of simple rules. Ranging from natural disasters such as earthquakes, forest fires and floods(Bak and Tang 1989; Malamud, Morein, and Turcotte 1998; Newman 2005) to human activities such as the distribution of city sizes, the frequency of words in language use, the wealth distribution, and the number of participants in strikes (Zipf 1949; Simon 1955; Newman 2005; Biggs 2005) the universal pattern called "power law" has been observed (Barabasi,2005;Lux,T et al,1999; Cederman et al.,2003; Gabaix et al,2003).

Among the works, the most celebrated work was the power law relationship of the frequency and severity of interstate wars and other conflicts. (Richardson,1994;Richardson,1998;Richardson,1960). Subsequent researches have proved that this relationship also holds for armed conflict issues; frequency and severity of conflicts (Roberts and Turcotte,1998;Cederman,2003), interstate and intrastate conflicts (Alvarez-Ramirez, Rodriguez and Urrea, 2007; Bohorquez et al., 2009; Scharpf et al., 2014; Trinn, 2015; Friedman, 2015), terrorist attacks (Clauset, Young and Gleditsch, 2007; Clauset, Woodard et al., 2013; Pinto, Lopes and Machado, 2012).

One of the salient characteristics of power law that distinguishes itself from a normal distribution is its "heavy-tail". A Power law distribution exhibits a nontrivial amount of weight far from its norm, which is not negligible.

This latent behavior, a power-law distribution, on the severity of conflicts implies that only the little portion of the events are severe with significant huge fatalities, while most of the events which account most of the frequencies are mere in degree.

The seminal point of the latent behavior of conflicts is that the sizes, or the severities of these incidents followed the precise statistical pattern, which is power-law distribution. The power-law is the probability of an incident that kills x people is $\Pr(x) \propto \mathrm{x}^{-\alpha}$, for all $x \geq x_{\min} > 0$ and where $\alpha > 1$ is called scaling parameter.

This formation shows that our expectations of a linear or normally distributed world may not hold for abounding cases. The average value, for example, is not a representative figure for the entire distribution, since the observation points above the mean is nontrivial.

There is a need to focus on the anomalies, the heavy-tail, or the severe events, since it is no more negligible, and it is the one which account for most of the severity. In the policy perspective, targeting the anomalies would be conducive in taking efficient preventive action. This research focuses on anomaly detection of irregular events, where the number of fatalities, or the severity of an event is huge.

2.4 Anomaly Detection

An anomaly or outlier is defined as a data point which differs significantly from the remaining data. The anomalies are observations that deviate from the remaining observations so much that induces it is suspected to be generated from a different mechanism (Hawkins,1980). Detecting the anomalies is an important task since it uncovers the characteristics of the data generation process. It is applied in credit card fraud detection(Brause,R.,et al,1999), health care system to detect unusual patient (Antonelli,D., et al.,2013), detecting erroneous treatment plans (Sipes, T. et al, 2014), other social issues, such as to detect anomalous increase in opioid deaths (Herlands, William, et al,2018), and increased 311 calls (Neill,Daniel B, 2018).

There are also previous literature of anomalous terrorist attacks or criminal detection. There are subtle differences by which methodology it uses, and which target it focuses on. Elovici conducted real-time analysis of potential profile of terrorists detecting web traffic using data mining techniques (Elovici et al. 2004). The other researches are detecting criminal behaviors using data mining techniques such as undersampling, oversampling techniques (Barbieris, 2014); Discerning the anomaly of terrorist attacks using Support Vector Machine (SVM), Naive Bayes (NB), and Logistic Regression (Meng, et al, 2017)

2.4.1 Anomaly Detection Methods

There are various means to detect the anomalies. The diverse methods can be chiefly classified into proximity based, probabilistic & statistical, and representational models. It differs in specific methods it adopts, however, the common concept is to implicitly generate a small summary of the data and the deviations from this summary are flagged as outliers. (Aggarwal,C.C.,2015).

Proximity-based method

Proximity-based methods assumes outliers as points isolated from the remaining data, based on the similarity or distance functions. It can be applied in one of three ways, nearest neighbor (distance based), density based, clustering methods. For nearest-neighbor method, the score of being an outlier comes from the distance of each data point to its kth nearest neighbors. For clustering methods, clustering algorithm is used to identify the dense regions and the data points fitness to the different regions functions as an outlier scores. Density-based method measures the sparseness of the data points, and the propensity of being outlier is discerned by the sparseness.

Probabilistic and statistical method

Probabilistic and statistical method models the data in the form of probability distribution and it learns the parameters from the distribution. For parametric models, Gaussian mixture models with Expectation-maximization (EM) algorithm is used to maximize the likelihood (or probability) of the observed data. Data points with the low fit values are considered outliers. Its advantage is that it can be applied to any type of data as long as there is a generative model for the data specification. However, a drawback is that certain probability distribution is assumed a priori, often engendering overfitting issue where outliers fit to the model. To overcome this problems, nonparametric models such as Gaussian Process, with kernel density estimation is used.

Representational methods

This method attempts to represent the observed data in terms of lower dimensional subspaces. Linear regression, for example, is a basic method, where the least-squares error value can be the outlier score. Principal component analysis (PCA), or spectral analysis is analogous in the concept of dimensionality reduction. The reconstruction error is used as an outlier score, where higher reconstruction errors indicate its behavior of deviation from latent representations. In this manner, autoencoders which aim to reconstruct the input as an output can be used as an anomaly detection model.

Chapter 3. Methodology and Data

3.1 Model

To illustrate the methodology of variational autoencoder (VAE), the concept of variational inference and the simple structure of autoencoder is explained beforehand. The concept of variational autoencoder is the confluence of variational inference and autoencoder with neural network.

3.1.1 Variational Inference

The basic concept of variational inference is to infer the posterior via the model Q_ϕ by optimizing the parameter ϕ so that it is 'close' to P. To calculate the closeness of the distribution, KL divergence is used. Eq 3-1 measures the difference of distributions of $P(Z)$ between $Q_\emptyset(Z)$. The parameter $\emptyset$ is optimized to minimize the KL divergence.

$$KL(Q_\phi(Z|X)||P(Z|X)) = \sum_{z \in Z} q_\phi(z|x) \log \frac{q_\phi(z|x)}{p(z|x)} \qquad \text{Eq 3-1}$$

Variational Autoencoder inferences approximate density via variational approximation. Ian Goodfellow classified the variational inference methods according to its specific tactics, and the Figure 1 shows the classification of variational inference. (Goodfellow,I.,2016)

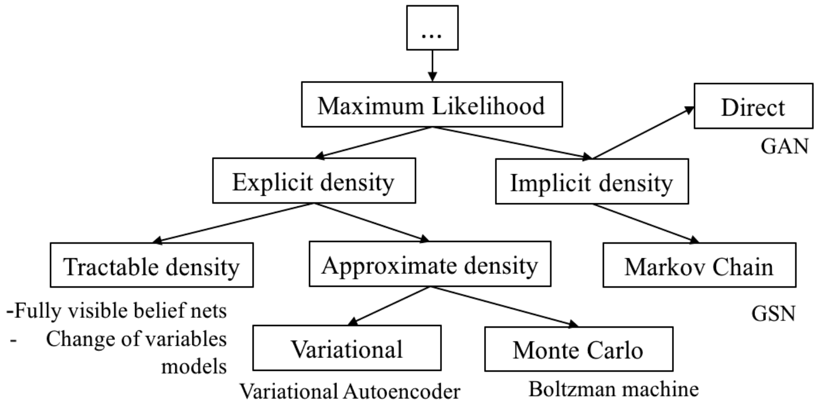

Figure 1 Variational Inference Method Classification

Variational Autoencoder is different from Markov Chain in that it infers explicit density, rather than implicit density. Implicit density methods don't explicitly define the model. Instead, it uses sampling as a method to extrapolate the true distribution. Given the sample x and transition operator q, repeatedly drawing sample x' is known to converge to the sample drawn from $p_{model}(x)$ under a certain assumption.

$$x' \sim q(x'|x)$$

Explicit density inferencing methods, on the other hand, attempts to define the model $p_{model}(x;\theta)$ and optimize it in the manner of maximizing the likelihood. By defining a model, it is much easier to handle while there exists a limitation which the model inherently encompasses.

Variational Autoencoder comes under the category of explicit density inference method. For inferencing the density in explicit density manner, it assumes the intractability, rather than tractability. The reason why the model approximates the true distribution can be confirmed from the Figure 2.

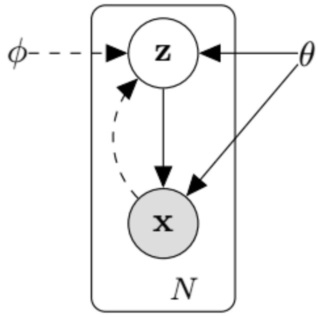

Figure 2 Directed Graphical Model

Figure 2 shows the basic directed graphical model. The observed sample data x is assumed to have a random process impact from the unobserved, latent variable z. The real line $p_\theta(z)p_\theta(x|z)$ represents the generative model. The dotted line $q_\phi(z|x)$ is used to approximate $p_\theta(z|x)$. To know the posterior $p_\theta(z|x) = p_\theta(z)\mathrm{p}_\theta(x|z)/p_\theta(\mathrm{x})$, there is a need to know the denominator $p_\theta(\mathrm{x})$=$\int_z p_\theta(x|z)p_\theta\,(z)dz$. However, as the z variable is latent and unknown, the posterior is mostly intractable. The intractability rises from the fact that the optimal model parameter θ^* and latent variable z is unknown.

To solve this intractability, the Varational Autoencoder uses variational approximation. It approximates the posterior via lower bounding technique. $L(x;\theta) \leq \log p_{model}(x;\theta)$

Since $\log p_{model}(x;\theta)$ is intractable, tractable lower bound is set. Maximizing the likelihood is now converted into maximizing the lower bound issue. The essence of variational approximation is that it converts statistical inference problem, which is to estimate the posterior, into an optimization problem.

3.1.2 Autoencoder

An autoencoder is a structure that aims to reconstruct itself to its input, a specific output or target label is absent, and it is trained in an unsupervised manner. It is normally implemented by neural networks, composed of two components, an encoder, and a decoder. A simplistic single hidden layer form of autoencoder can be denoted as Eq 3-2 and Eq 3-3.

$$h = \sigma(W_{xh}x + b_{xh}) \tag{Eq 3-2}$$

$$x' = \sigma(W_{hx}h + b_{hx}) \tag{Eq 3-3}$$

$$\|x - x'\| \tag{Eq 3-4}$$

The W and b respectively is the weight and bias, and σ is the affine transformation function, which gives nonlinearity of the representation. x is the input vector, h is the hidden representation layer. Eq 3-2 is the encoder function, where the input vector is mapped into hidden layer h, an affine transformation. Eq 3-3 is the decoder function, where the hidden layer h is mapped to the original input vector x. Eq 3-4 represents the reconstruction error of the original input x and the reconstruction output x'. An autoencoder optimizes its parameters to minimize this reconstruction error.

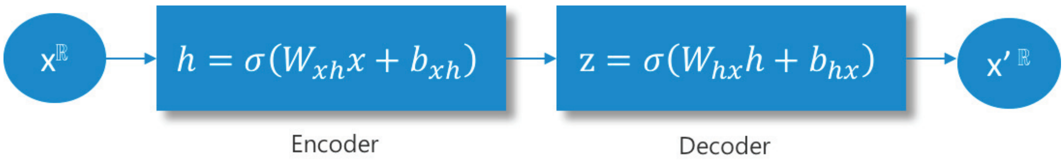

Figure 3 The basic Autoencoder structure

3.1.3 Variational Autoencoder (VAE)

The term ‘variational’ implies the variational approximation inference it uses and the term ‘autoencoder’ indicates the autoencoder structure. From the Figure 2, the generative model $p_\theta(z)p_\theta(x|z)$, leaving the prior knowledge $p_\theta(z)$ aside, $p_\theta(x|z)$ indicates to the model that generates the sample x given the latent variable z. The hidden layer h of the autoencoder and the latent variable z construed as the representation layer of autoencoder, $p_\theta(x|z)$ corresponds to the decoder, which generates the sample from the given hidden layer. $q_\phi(z|x)$, on the other hand, corresponds to the encoder where gives the code given the sample x.

The marginal likelihood $\log p_\theta$ is defined by the sum of marginal likelihood of each individual data point. The individual marginal likelihood of each data point is shown in Eq 3-5.

$$\log p_\theta\left(x^{(1)}, \ldots, x^{(N)}\right) = \sum_{i=1}^{N} \log p_\theta\left(x^{(i)}\right) \qquad \text{Eq 3-5}$$

$$\begin{aligned}\log p_\theta\left(x^{(i)}\right) &= D_{KL}(q_\emptyset(z|x^{(i)})||p_\theta\left(z|x^{(i)}\right)) \\ &\quad + \mathbb{E}_{q_\emptyset(z|x)}[-\log q_\emptyset(z|x) + \log p_\theta(x,z)]\end{aligned} \qquad \text{Eq 3-6}$$

The first term in Eq 3-6 indicates the difference of the true posterior distribution and the approximated distribution (KL divergence) which is always larger than 0. The second term is the variational lower bound. Eq 3-6 can be put into another form, as Eq 3-7, Eq 3-8

$$\log p_\theta\left(x^{(i)}\right) \geq \mathbb{E}_{q_\emptyset(z|x)}[-\log q_\emptyset(z|x) + \log p_\theta(x,z)] = \mathcal{L}(\theta, \emptyset; x^{(i)}) \qquad \text{Eq 3-7}$$

$$\mathcal{L}\big(\theta,\emptyset;x^{(i)}\big) = -D_{KL}(q_{\emptyset}(z|x^{(i)})||p_{\theta}(z)\big) + \mathbb{E}_{q_{\emptyset}(z|x)}[\log p_{\theta}(x^{i}|z)] \qquad \text{Eq 3-8}$$

The Eq 3-8 , indicates by optimizing the parameters $\theta,\emptyset$, it aims to maximize the lower bound $\mathcal{L}$. This equation is simplified to Eq 3-9.

$$(\theta^{*},\phi^{*}) = \arg\max_{\theta,\phi}\mathcal{L}\,(\theta,\phi;x^{(i)}) \qquad \text{Eq 3-9}$$

Solving Eq 3-9 can be interpreted as solving the maximum likelihood $\log p_{\theta}(x^{i}|z)$, by minimizing the difference of the variational approximation $q_{\phi}(z|x^{i})$ and the prior of Z, which is $p_{\theta}(z)$, by adding a regularization term.

Variational autoencoder approximate the posterior by using neural network. One of the powerful tools of neural network is to use stochastic gradient descent or ascent in feedforward and back-propagation to optimize parameters. To utilize the neural network, it uses the re-parameterization trick. According to the Eq 3-8, $\mathbb{E}_{q_{\emptyset}(z|x)}[\log p_{\theta}(x^{i}|z)]$ samples from q, $z{\sim}q_{\phi}(z|x)$. The feed-forward works by sampling z from the distribution $q_{\phi}(z|x)$, and calculating $p_{\theta}(x|z)$ from the sampled z. However, since sampling is not a differentiable operation, there is a difficulty when applying backward-propagation. To update the stochastic gradient it requires a differentiable parameter. For the fixed parameters, the stochasticity is in the input and from the given input, the output is always the same. However, sampling gives the stochasticity in the model itself. Therefore, to make use of neural network and to allow the backward propagation, it uses re-parameterization trick.

$$\tilde{z} = g_{\phi}(\epsilon, x)\ with\ \in \sim p(\epsilon) \qquad \text{Eq 3-10}$$

The conversion differentiable function for the re-parameterization $g_{\phi}(\epsilon, x)$, with the noise variable $\in$, is shown in Eq 3-10. The Monte Carlo expectation estimate of random function $f(z)$ for $q_{\phi}(z|x)$ is like Eq 3-11.

$$\mathbb{E}_{q_{\phi}\left(z\middle|x^{i}\right)}[f(z)] = \mathbb{E}_{p(\epsilon)}\left[f\left(g_{\phi}(\epsilon, x^{i})\right)\right]$$

$$= \frac{1}{L}\sum_{l=1}^{L} f\left(g_{\phi}(\epsilon^{(l)}, x^{(i)})\right), \quad where\ \epsilon^{(l)} \sim p(\epsilon) \qquad \text{Eq 3-11}$$

$$\tilde{\mathcal{L}}\left(\theta, \phi; x^{(i)}\right) = -D_{KL}(q_{\phi}\left(z\middle|x^{(i)}\middle\|p_{\theta}(z)\right)$$

$$+\frac{1}{L}\sum_{l=1}^{L}(\log p_{\theta}(x^{(i)}|z^{(i,l)}))\,, where\ z^{(i,l)} \qquad \text{Eq 3-12}$$

$$= g_{\phi}\left(\epsilon^{(i,l)}, x^{(i)}\right)\ and\ \epsilon^{(l)} \sim p(\epsilon)$$

The single value estimate version of Eq 3-11is shown in Eq 3-12. Unlike the previous sampling operation $z \sim q_{\phi}(z|x)$, which the model itself had stochasticity, has changed to $g_{\phi}\left(\epsilon^{(i,l)}, x^{(i)}\right)\ and\ \epsilon^{(l)} \sim p(\epsilon)$. Table 1 shows the VAE training algorithm.

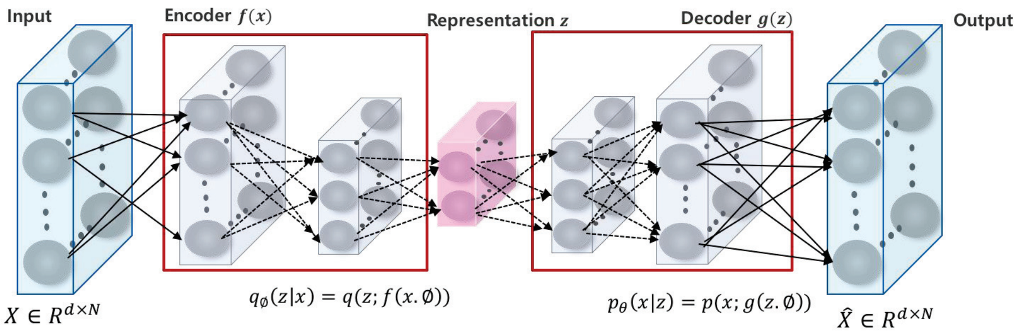

Figure 4 The Figure of VAE

VAE Training algorithm

INPUT: Dataset $x^{(1)},..,x^{(N)}$

OUTPUT: encoder f_{ϕ}, decoder g_{θ}

$\phi, \theta \leftarrow$ Initialize parameters

repeat

for i=1 to N do

Draw L samples from $\epsilon \sim N(0,1)$

$z^{(i,l)} = g_{\phi}\left(\epsilon^{(i)}, x^{(i)}\right) \quad i = 1, \dots, N$

end for

$$E = \sum_{i=1}^{N} -D_{KL}(q_{\phi}(z|x^{(i)})||p_{\theta}(z)) + \frac{1}{L}\sum_{l=1}^{L}(\log p_{\theta}(x^{(i)}|z^{(i,l)}))$$

ϕ, θ $\leftarrow$ Update parameters using gradients of $\boldsymbol{E}$ (e.g. Stochastic Gradient Descent)

until ϕ, θ converge

Table 1 VAE Training Algorithm

Figure 4 shows the stacked hidden layer variational autoencoder. Stacking the layers, it forms deep autoencoders. Reducing the dimension in a hierarchical manner, the performance, which is to extract relevant features and to marginalize the noises, the stacked version of VAE is known to have enhanced ability. The number of hidden units in the hidden layers diminishes to avoid lookup table like structure.

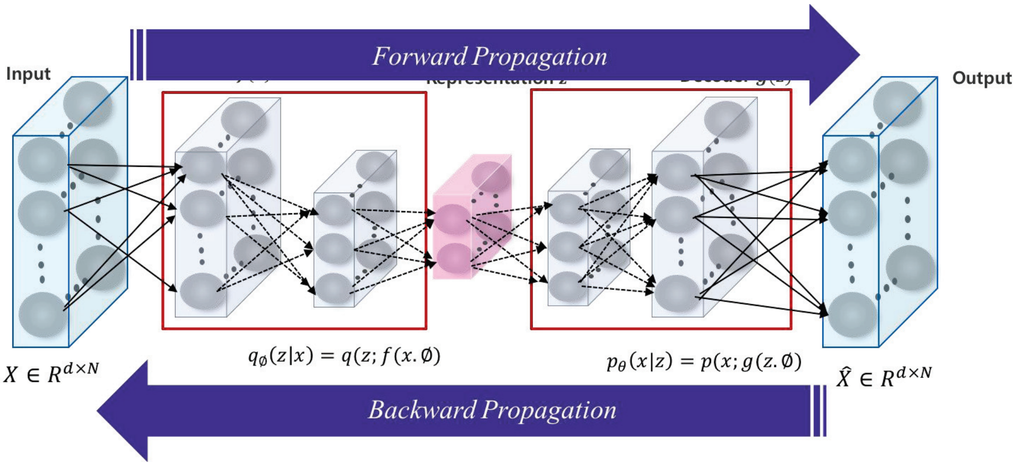

Figure 5 Forward and Back Propagation of VAE

3.2 VAE Anomaly detection

Table 2 shows the anomaly detection algorithm. The reconstruction error probability shows the fallacy rate of reconstructing itself. Utilizing the structure of VAE autoencoder as anomaly detector, the reconstruction error itself can be interpreted as probability of being anomaly. Higher reconstruction error means higher possibility of being anomaly since its characteristics deviates from the characteristics or dynamics trained from the trained model, which is trained from only normal events.

VAE Anomaly Detection algorithm

INPUT: Training dataset X, X* (where X_i labeled Normal) ,

Test dataset $x^{(i)}$ (i=1..n)

OUTPUT: reconstruction error rate $|x - \hat{x}|$

$\phi, \theta \leftarrow$ Initialize parameters

for i=1 to *N* do

μ_{z^i} , $\sigma_{z^i} = f_\theta(z|x^i)$

for l=1 to L do

$\mu_{\hat{x}^{(i,l)}}$, $\sigma_{\hat{x}^{(i,l)}} = g_\emptyset(x|z^{i,l})$

end for

reconstruction error rate(i) = $\left|x - \frac{1}{L}\sum_{l=1}^{L} p_\theta\left(x^{(i)}\middle|\mu_{x^{(i,l)}} , \sigma_{x^{(i,l)}}\right)\right|$

end if

end for

Table 2 VAE Anomaly Detection algorithm

3.3 Analysis Sequence

Figure 6 shows the process of analysis. Firstly, it goes through the preprocessing. The data is filtered by country and attack type. Attributes extracted is latitude, longitude, time and the number of fatalities.

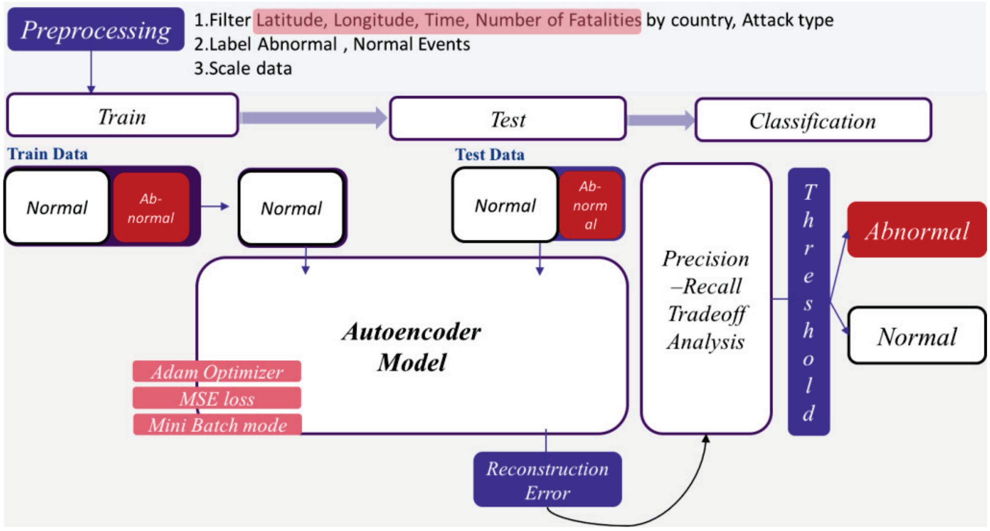

Figure 6 Analysis Sequence

Using the number of fatalities, the label is made to discern the normal and the abnormal event. It is trained in semi-supervised manner, where the label is used indirectly rather than direct way for training. The measure of classifying normal and abnormal events is explained in section 3.4. The data extracted have been scaled by minmax scaler. Next, the model is trained. Only the normal events are extracted and used to train the model. The autoencoder model uses Adam optimizer, mean square error loss to optimize the parameters. After the training is done, the model is tested by the test data. The test data consists of the normal and abnormal events. The output is the reconstruction error. Lastly, the reconstruction error is used to analyze the precision-recall tradeoff.

3.4 Data Description

The empirical analysis is based on data from the Armed Conflict Location & Event Data (ACLED). ACLED publicly provides armed conflict data covering various continents, diverse attack types, for a certain period of accumulated time. The continent and time period available are likewise: 1997~2019 for Africa, 2016~2019 for the Middle East, 2013~2019 for Asia, 2016~2019 for Europe. The classification of attack types can be found in Table 3. For the analysis, only the event type 'Battles' and 'Violence against civilians' among the 'violent events' has been used. The Analysis was conducted on Event-type level, where Sub-Event Type has been aggregated to event-type level. For each incident, the data reports the exact date (Year-Month-Day), geographical data (to Longitude, Latitude level), actors, interaction type, number of fatalities. Among these attributes, 'date', geo data: 'longitude', 'latitude', and 'number of fatalities' has been extracted for analysis.

Compared to other sources, such as GTD (Global Terrorism Database) or Icasualties (icasualities.org), ACLED provides data in a more specific and disaggregated manner. The specialty of this dataset is that the geo data is given with geo-precision, geographically disaggregated to the most detailed 'town' level. (GTD only provides city level precision) Another specialty of this data is that it affords disaggregation of events by its event type, which allows to observe the behavior and analyze it in the more sophisticated matter.

General	Event Type	Sub-Event Type
Violent events	Battles	Armed clash
		Government regains territory
		Non-state actor overtakes territory
	Violence against civilians	Sexual violence
		Attack
		Abduction/Forced disappearance

Table 3 Classification of the Violent events

The empirical analysis was conducted based on Syria. Syria was reported to have experienced one of the deadliest conflicts. Syria comprised 29% of the world's conflict-related deaths in 2017. Syria had the most lethal non-state conflict-related deaths in 2017 (Dupuy,K et al,2018).

The attributes used for analysis is the geo-data, latitude and longitude, temporal information, and the number of fatalities. The date of year, month, and day is converted into elapsed time based on the initial date. The events with any missing attribute values are dropped. The events with no fatalities are also dropped, and only the events with at least one minimum fatality are counted.

	Values
Number of events	Total 6042
Date	2017.04.01~2018.10.31
Latitude	32.424~37.186 (955 Latitude values)
Longitude	35.785~42.177 (944 Longitude values)
[Latitude, Longitude]	1161 values max 190 incidences per one site min 1 incidences per one site
Fatalities	Min 1 (1526 events), Max 469 (1 event)
Power law alpha	3.5525
Power law sigma	0.1987

Table 4 Data description

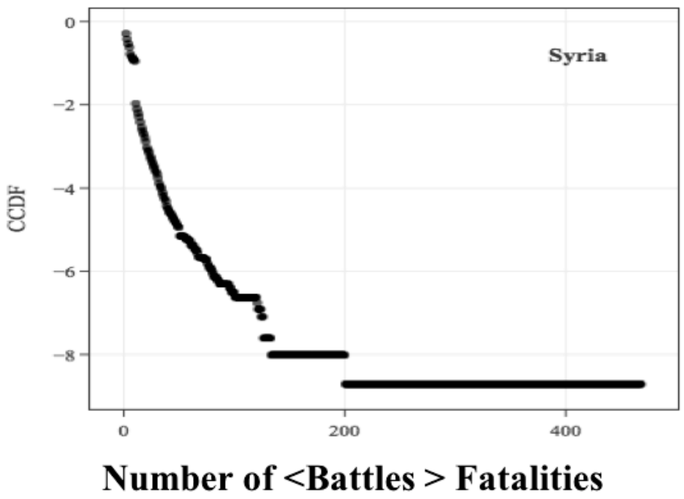

Figure 7 Complementary cumulative distribution function (CCDF) of Battle fatalities

Figure 7 shows the log-log complementary cumulative distribution function of Syria battle fatalities. According to the power-law analysis, the power law alpha is 3.5525, $\Pr(x) \propto x^{-\alpha}, \alpha \simeq 3.5525$. It conforms the previous literature of latent behavior of human insurgency, which is exhibiting the 'heavy tail'. This shows that only the little portion of the events are severe with significant huge fatalities, while most of the events which account most of the frequencies are mere in degree. There is a need to focus on the severe events with high fatalities, which account only the little portion among the total events. Such events are defined as abnormal events and it is defined where $p\left(x^{(i)}\right) \leq 0.02$

Figure 8 shows the probability distribution function of number of fatalities. The red line, which is probability of 0.02, classifies the normal and abnormal events. Among total 6042 events, abnormal events were 121 events. 20% of the data, which is 1208 events were separated as test data. Figure 9 is the plotted version of normal and abnormal events.

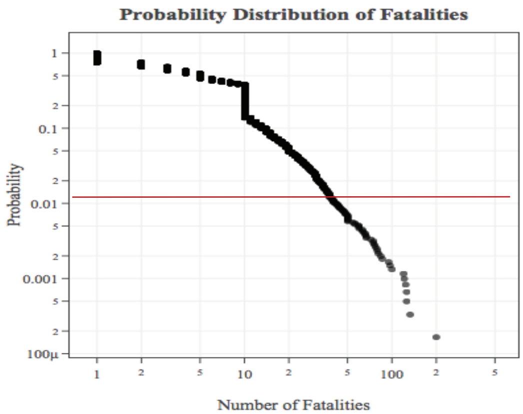

Figure 8 Probability Distribution of Fatalities

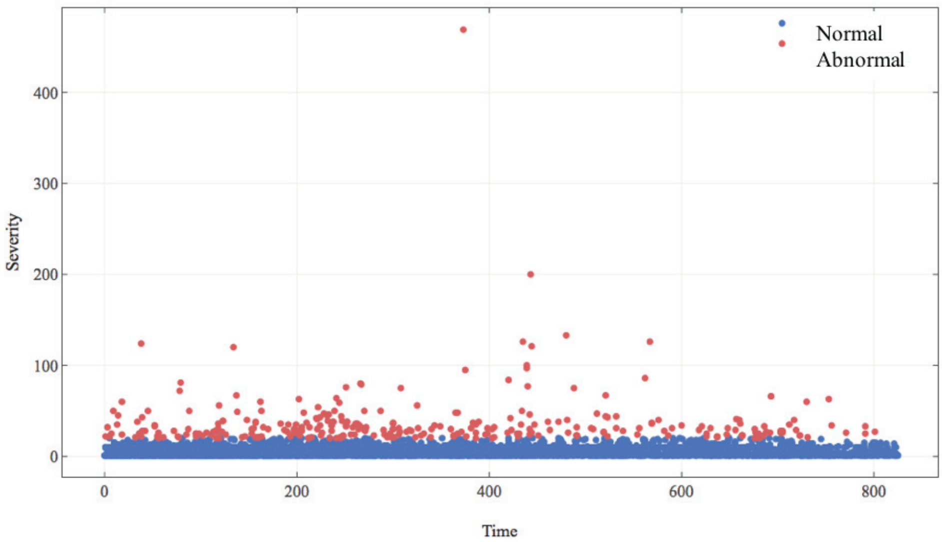

Figure 9 Plot of Normal and Abnormal Classification

Chapter 4. Result

4.1 Reconstruction Error

The reconstruction error of the test result is shown in Figure 10. The blue dots and orange dots represent the original classification that was made before the test. The overall trend shows that the abnormal events have high reconstruction error with high variance while the normal events' reconstruction error show little variance with minimal reconstruction error.

Figure 11 is the detailed histogram of the reconstruction error of normal events and abnormal events respectively. The range of reconstruction error of normal events is at most 0.25, while that of abnormal events is at most 5. Higher reconstruction error indicates higher possibility of being anomaly since its characteristics deviates from the characteristics or dynamics trained from the trained data, which is the normal events.

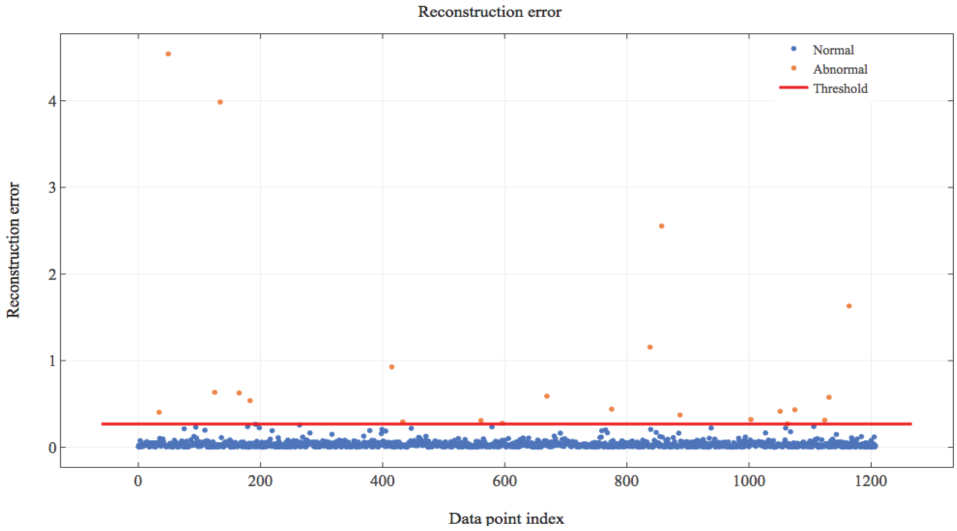

Figure 10 Result of Reconstruction Error

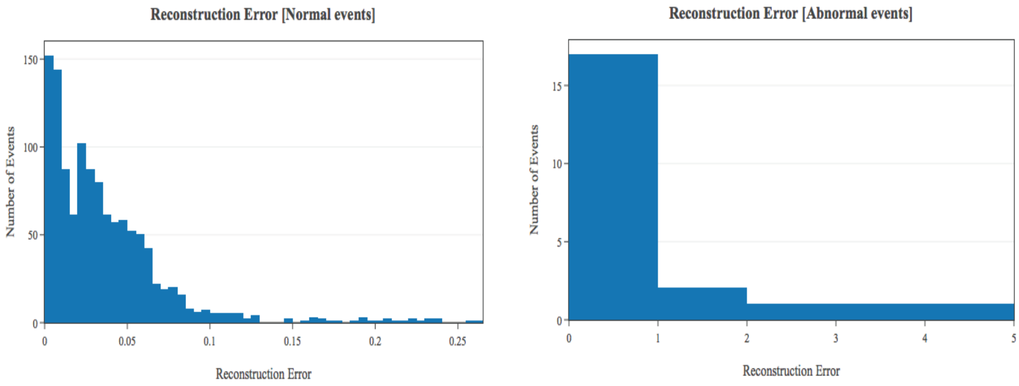

Figure 11 Reconstruction Error of Normal and Abnormal events

Figure 12 shows the mean reconstruction error of anomaly predicted test points. It provides an attribute analysis for which attribute is contributing to the arise of abnormal decision. The 'latitude' value is 0.08, 'longitude' is '0.05' and 'time' is 0.18. It can be construed that 'time' component is the most considerable attribute, that increases the propensity of being anomalous event, which reflects the abrupt and unpredictable nature of attacks. Spatial component of 'Latitude' and 'Longitude', compared to the temporal component, is less significant in increasing the propensity of engendering abnormal event. This can be construed as the location where abnormal events happen, is relatively predictable or static than the temporal component.

Figure 12 Reconstruction Error of Attributes

4.2 Performance Analysis

The measure of performance is followed likewise: Accuracy, ROC, precise and recall value. This measure is defined by the combination of the classification of Table 5. The true value is the actual classification and the predicted value is the result made by test process. 'True' is the anomaly event, 'False' is the normal event and 'Positive' is the anomaly predicted event, 'Negative' is the normal predicted event. 'True Positive' is the abnormal

		True Value	
		True	False
Predicted Value	Positive	True Positive	False Positive (Type I error)
	Negative	False Negative (Type III error)	True Negative

Table 5 Categorization of classification

event predicted as abnormal event. 'False Positive' is the Type I error where normal event is predicted as abnormal. 'False Negative' is the Type III error where abnormal event predicted as normal event. 'True Negative' is the case where normal event predicted as normal event.

4.2.1 Accuracy

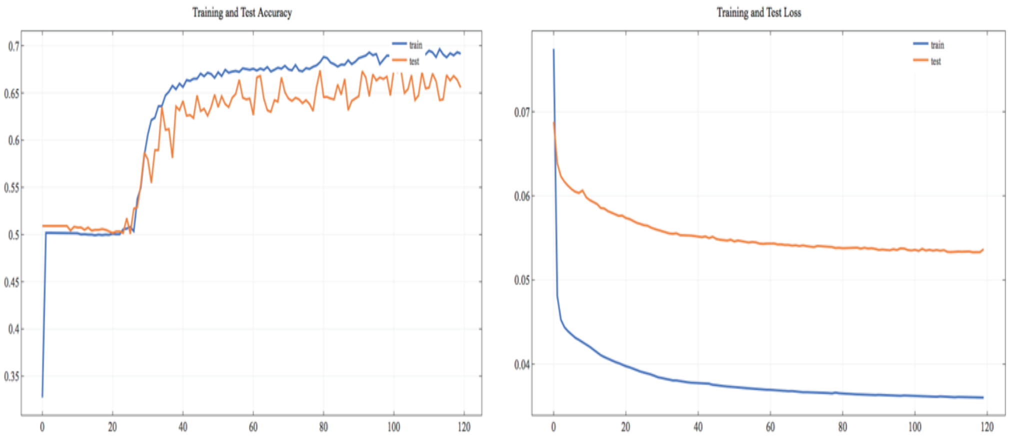

Figure 13 The accuracy and the loss

Accuracy measure is defined as Eq 4-1. The accuracy is the ratio of true positive and true negative among all the categorization combined. Figure 13 shows the train and test accuracy and loss respectively. Accuracy is one of the most widely applied measure in performance analysis of the model. This task, however, the error is deliberately made so the accuracy measure is not apt for analyzing the performance of the model.

$$Accuracy = \frac{True\ Positive + True\ Negative}{TP + FP + TN + FN} \qquad \text{Eq 4-1}$$

4.2.2 ROC (Receiver Operating Characteristics)

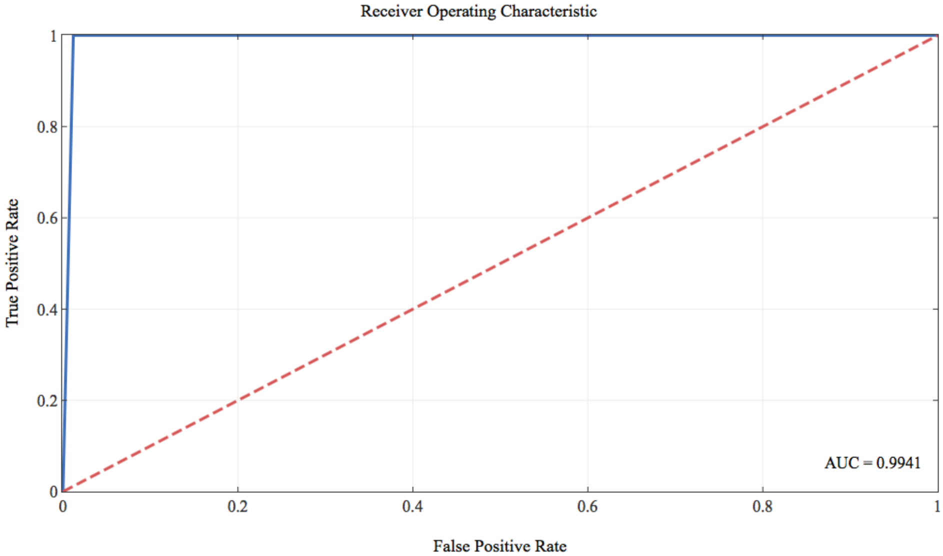

Figure 14 ROC curve

ROC curve shows the true positive rates by false positive rate. Blue line in Figure 14 shows the ROC curve. Blue line as close to the upper left corner, means that it can correctly discern True Negative and True Positive events. The ROC curve of this case implies that it can correctly discern the true negative and the true positive events. However, there is a need to check that the data used for the analysis is highly lopsided, where only 2% of the data are classified as the anomalies. Therefore, the curve itself cannot be the measure for the model and other measures should be taken into account.

$$True\ Positive\ Rate = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad \text{Eq 4-2}$$

$$False\ Positive\ Rate = \frac{False\ Positive}{True\ Negative + False\ Positive} \quad \text{Eq 4-3}$$

4.2.3 Precision

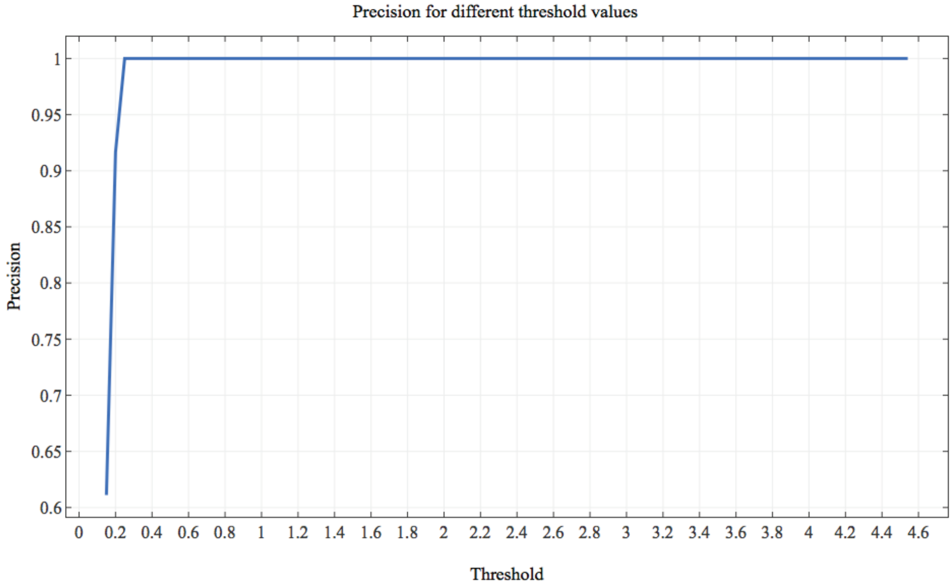

Figure 15 Precision by threshold

Precision measures how many selected events are relevant. It measures the relevancy of obtained results. The relevancy is to not predict a negative sample as positive. If precision equals 1, then it indicates False Positive case, normal events predicted as abnormal events, doesn't exist. Eq 4-4 is the definition of the precision.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad \text{Eq 4-4}$$

Figure 15 presents the precision value by threshold. It can be seen from Figure 15 that as threshold increases, precision value is constant with 1. However, if the threshold is lower than 0.2 beyond, precision value is in decreasing trend, which indicate that probability of normal events predicted as abnormal events increases.

4.2.4 **Recall (Sensitivity)**

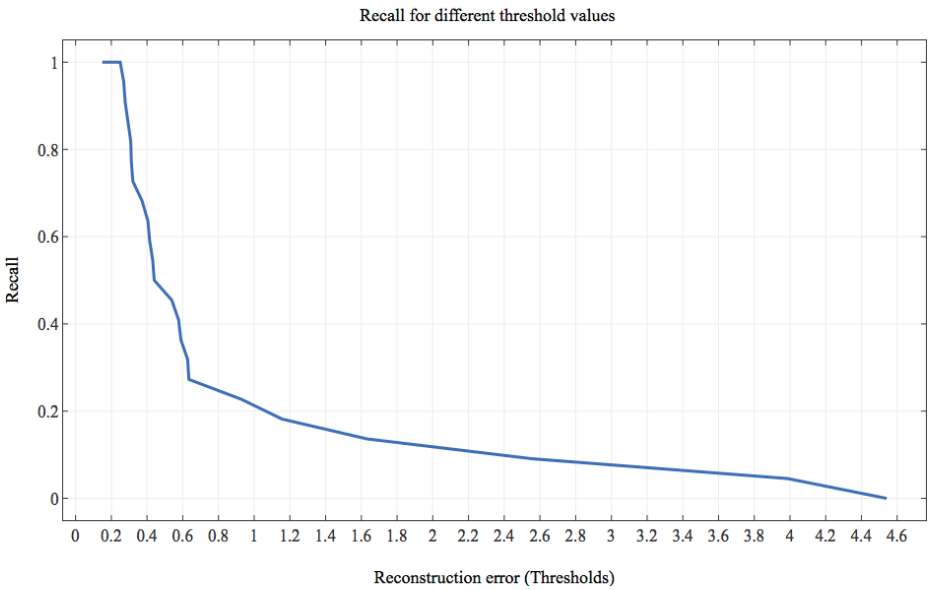

Figure 16 Recall by Threshold

Recall (Sensitivity) measure shows how many relevant events are selected. It measures the ability to find all the positive samples. If recall equals 1, it indicates that False Negative, abnormal events predicted as normal events doesn't exist.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad \text{Eq 4-5}$$

Figure 16 shows the change of recall value based on the threshold value. As the threshold increases, the recall rate decreases, which indicates that less relevant events—abnormal events predicted as normal events— are selected.

4.2.5 Precision vs Recall

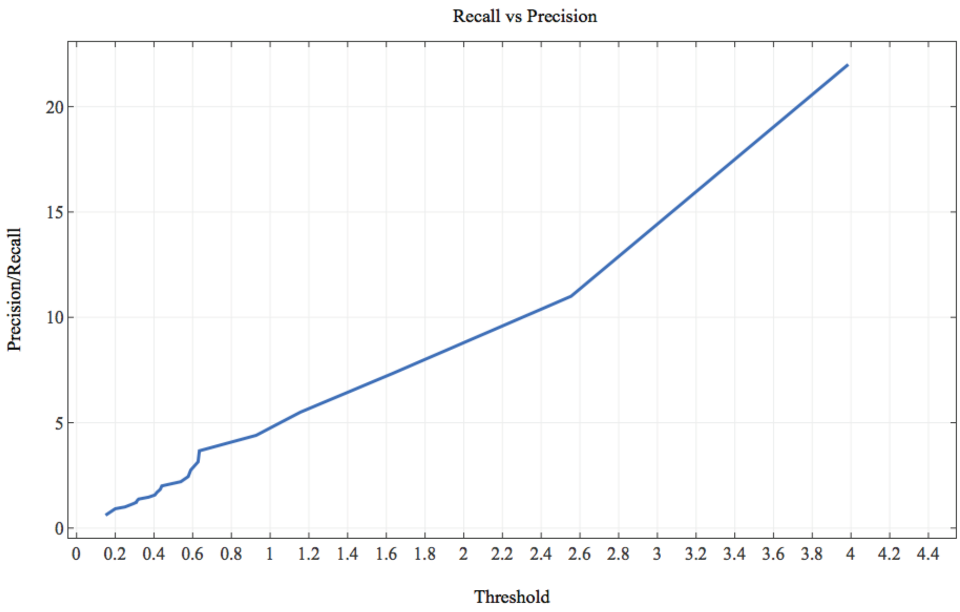

Figure 17 Precision/Recall by threshold

The previous analysis of precision and recall shows that there is a trade-off between the precision and recall value. Figure 17 and Figure 18 show in detail how the trade-off between precision and recall value differs within threshold. Figure 17, the precision/recall value increases as the threshold value increases. This means that the precision value changes in an increasing manner while recall value decreases. While the trade-off between the recall and precision rate exists, the threshold that minimizes the trade-off which guarantees the highest value of both the precision rate and recall rate is chosen. From the precision by recall graph in Figure 18, the threshold which satisfies the highest value of precision rate which equals to 1 and the recall rate which also equals to 1 is chosen. The value of threshold was 0.269.

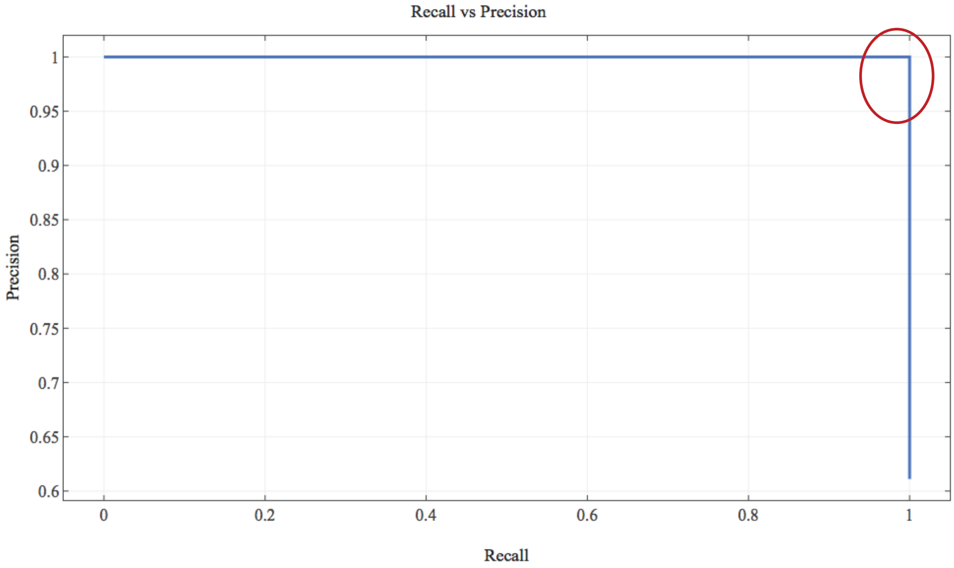

Figure 18 Precision by Recall

Table 6 shows the result of classification. From the threshold, it successfully classified the events abnormal as abnormal, normal as normal, showing no Type I or Type III error.

Confusion Matrix		*Predicted Class*	
		Normal	Abnormal
True Class	Normal	1186	0
	Abnormal	0	22

Table 6 Confusion Matrix

Chapter 5. Discussion and Conclusion

5.1 Implication

This research focuses on detecting the anomalies. The result implies that the VAE model can successfully detect the normal and abnormal events given the space-temporal information. There has been fairly good academic work applied to the real policymaking world (Metternich et al., 2013; Brandt, Colaresi & Freeman, 2008). This model could also give implication to understand and model violence. Especially, for the UN Troop deployment for preventing violence against civilians (Phayal.A,2019) and for government's deploying its forces for preventing human insurgencies, this space-temporal model could be conducive for giving temporal and spatial information for the prospective deployment.

The structure of VAE enables to analyze which attributes are contributing to the outbreak of the anomalies — severe events. The Syria battle case shows that the temporal component account most of its anomaly outbreak, which reflect the abruptness behavior of these anomalies.

Also, the VAE methodology gives the objective probability, where threshold is not fixed and through the precision-recall analysis, optimal threshold can be derived. Choosing an optimal threshold is an important matter in real policy making, which functions as a criteria or indicator for allocating its resources.

5.2 Limitations

Notwithstanding its contribution, there still exists limitations. Firstly, the data used for analysis is dependent on single source, ACLED. Although ACLED itself collaborates the information from diverse media sources and local information, the inherent bias that the data essentially possess, could not be disregarded. For the data reliability, the geo data of longitude and latitude accuracy, though the precision is refined, cannot be guaranteed. Secondly, the model for analysis is tested only for specific country, certain attack type, which means that the data it uses is disaggregated by country, and attack type. Since the dynamics of conflict differs by country, and the type of attacks, the model is tested specifically in the disaggregated level. This disaggregated analysis of conflict renders a fine-grained result, permitting a more nuanced analysis of conflicts (Donnay,K et al, 2014). Its analysis accounts for changes over time and across spatial units in the incidence, intensity and duration of events. However, such level analysis involves the trade-off in sacrificing the greater external validity for internal validity. The variation of subnational level dismisses the cross-national studies which yield broadly applicable findings (Donnay,K et al, 2014). Also, there is no consensus of the appropriate level of disaggregation. The questions to be further dealt with is likewise: 'In what means different datasets on conflict be linked to each other and to data on other factors? How can challenges be addressed by using disaggregated data? What are the weaknesses and the strengths of this statistical analysis?'

5.3 Further Research

This research focuses on the basic but the most essential component, spatial and temporal information. For more specific and detailed analysis, analysis with other covariates would be also plausible. Analysis by actors, for example, would also incorporate the heterogeneous characteristic of each actors, and uncover the variation of its propensity to engage in violence (Bhavnani & Backer 2000; Humphreys & Weinstein 2008; Verwimp 2006). Other socio-political indicators are also known to affect conflicts, such as large-scale population shifts, changes of political structures.

Bibliography

ACLED. (2019). “Armed Conflict Location & Event Data Project (ACLED) Codebook, 2019

Adelaja, A., & George, J. (2019). Effects of conflict on agriculture: Evidence from the Boko Haram insurgency.*World Development*,*117*, 184-195.

Aggarwal, C. C. (2015). Outlier analysis. In*Data mining* (pp. 237-263). Springer, Cham.

Alvarez-Ramirez, J., Rodriguez, E., & Urrea, R. (2007). Scale invariance in the 2003–2005 Iraq conflict. *Physica A: Statistical Mechanics and its Applications*,*377*(1), 291-301.

Antonelli, D., Bruno, G., & Chiusano, S. (2013). Anomaly detection in medical treatment to discover unusual patient management. *IIE Transactions on Healthcare Systems Engineering*,*3*(2), 69-77.

Bak, Per, and Chao Tang. (1989). Earthquakes as a self-organized critical phenomena. Journal of Geophysical Research 94 (B11): 15635-7.

Barabasi, A. L. (2005). The origin of bursts and heavy tails in human dynamics.*Nature*,*435*(7039), 207.

Bhavnani, R. (2006). Ethnic Norms and Interethnic Violence: Accounting for Mass Participation in the Rwandan Genocide. Journal of Conflict Resolution, 43 (6), 651–69.

Biggs, Michael. 2005. Strikes as forest fires: Chicago and Paris in the late 19th century. American Journal of Sociology 111 (1): 1684-714.

Bohorquez, J. C., Gourley, S., Dixon, A. R., Spagat, M., & Johnson, N. F. (2009). Common ecology quantifies human insurgency. *Nature*,*462*(7275), 911.

Boone, C. (2003). *Political topographies of the African state: Territorial authority and institutional choice*. Cambridge University Press.

Brandt Patrick T, Freeman John R, Schrodt Philip A (2011) Real time, time series forecasting of inter- and intra-state political conflict. Conflict Management and Peace Science 28(1): 41–64

Brandt Patrick, Colaresi Michael, Freeman John (2008) The dynamics of reciprocity, accountability and credibility. Journal of Conflict Resolution 52(2): 343–374.

Brause, R., Langsdorf, T., & Hepp, M. (1999). Neural data mining for credit card fraud detection. In *Proceedings 11th International Conference on Tools with Artificial Intelligence* (pp. 103-106). IEEE.

Breckner, M. (2019).*Climatic and geographic determinants of economic development* (Doctoral dissertation, lmu).

Bueno de Mesquita, Bruce (2010). The Predictioneer's Game: Using the Logic of Brazen Self-Interest to See and Shape the Future.New York:Random House.

Cederman, L. E. (2003). Modeling the size of wars: from billiard balls to sandpiles.*American Political science review*,*97*(1), 135-150.

Clauset A, Gleditsch KS (2012), "The Developmental Dynamics of Terrorist Organizations",PLOS ONE,7(11): e48633,doi:10.1371/journal.pone.0048633,PMC3504060,PMID23185267

Clauset, A., Young, M., & Gleditsch, K. S. (2007). On the frequency of severe terrorist events.*Journal of Conflict Resolution*,*51*(1), 58-87

De-Arteaga, M., Herlands, W., Neill, D. B., & Dubrawski, A. (2018). Machine learning for the developing world. *ACM Transactions on Management Information Systems (TMIS)*,*9*(2), 9.

Donnay, K., Gadjanova, E., & Bhavnani, R. (2014).*Disaggregating conflict by actors, time, and location*(pp. 44-56). Paradigm.

Douglas M Hawkins. Identification of outliers, volume 11. Springer, 1980

Dupuy, K., & Rustad, S. A. (2018). Trends in Armed Conflict, 1946–2017.*Conflict Trends*,*5*, 2018.

Eck, J., Chainey, S., Cameron, J., & Wilson, R. (2005). Mapping crime: Understanding hotspots.

Elovici, Y., Kandel, A., Last, M., Shapira, B., & Zaafrany, O. (2004). Using data mining techniques for detecting terror-related activities on the web.*Journal of Information Warfare*,*3*(1), 17-28.

Federal Bureau of Investigation, Counterterrorism Threat Assessment and Warning Unit, Counterterrorism Division. 1999. Terrorism in the United States, 1999. US Department of Justice, Washington, DC. http://www.fbi.gov/publications/terror/terror99.pdf

Flaxman, S. R. (2014). A general approach to prediction and forecasting crime rates with Gaussian processes.*Heinz College Second Paper. Pittsburg: Carnegie Mellon*

University.

Friedman, E., & McCormick, B. L. (2015).*What if China doesn't democratize?: implications for war and peace*. Routledge.

Gabaix, X., Gopikrishnan, P., Plerou, V., & Stanley, H. E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature*,*423*(6937), 267.

Goodfellow, I. (2016). NIPS 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*.

Gorr, W., Olligschlaeger, A., & Thompson, Y. (2003). Short-term forecasting of crime.*International Journal of Forecasting*,*19*(4), 579-594.

Hegre, H., Metternich, N. W., Nygård, H. M., & Wucherpfennig, J. (2017). Introduction: Forecasting in peace research.

Höglund, K., Melander, E., Sollenberg, M., & Sundberg, R. (2016). Armed Conflict and Space: Exploring Urban-Rural Patterns of Violence. In Spatializing Peace and Conflict (pp. 60-76). Palgrave Macmillan, London.

Humphreys, M., & Weinstein, J. M. (2008). Who Fights? The Determinants of Participa-tion in Civil War. American Journal of Political Science, 52 (2), 436–55.

Jackson, P. (2019). Intelligence in a modern insurgency: the case of the Maoist insurgency in Nepal. *Intelligence and National Security*, 1-15.

Kalyvas, S. N. (2006).*The logic of violence in civil war*. Cambridge University Press.

Kim, H. (2019). In the wake of conflict: the long-term effect on child nutrition in Uganda.*Oxford Development Studies*, 1-20.

Kupilik, M., & Witmer, F. (2018). Spatio-temporal violent event prediction using Gaussian process regression.*Journal of Computational Social Science*,*1*(2), 437-451

Malamud, Bruce D., Gleb Morein, and Donald L. Turcotte. (1998). Forest fires: An example of selforganized critical behavior. Science 281:1840-1.

Martin-Shields, C. P., & Stojetz, W. (2019). Food security and conflict: Empirical challenges and future opportunities for research and policy making on food security and conflict. World Development,119, 150-164.

Meng, X., Mo, H., Zhao, S., & Li, J. (2017, April). Application of anomaly detection for detecting anomalous records of terroris attacks. In 2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA) (pp. 70-75). IEEE.

Metternich, Nils W, Dorff, Cassy, Gallop, Max, Weschle, Simon, Ward, Michael D (2013)Antigovernment networks in civil conflicts: How network structures affect conflictual behavior. American Journal of Political Science 57(4):892–911.

Morris, M. F. (2005).*al-Qaeda as Insurgency*. ARMY WAR COLL CARLISLE BARRACKS PA.

Nandwani, B. (2019). Decentralisation, Economic Inequality and Insurgency.The Journal of Development Studies,55(7), 1379-1397.

Neill, D. B., & Herlands, W. (2018). Machine learning for drug overdose surveillance.*Journal of Technology in Human Services*, *36*(1), 8-14.

Newman, Mark. E. J. (2005). Power laws, Pareto distributions and Zipf’s law.

Contemporary Physics 46 (5): 323-51.

O'Brien, Sean P (2002). Anticipating the good, the bad, and the ugly: An early warning approach to conflict and instability analysis. Journal of Conflict Resolution 46(6):791–811.

Odozi, J. C., & Oyelere, R. U. (2019).*Conflict Exposure and Economic Welfare in Nigeria*(No. 334). GLO Discussion Paper.

Phayal, A. (2019). UN Troop Deployment and Preventing Violence Against Civilians in Darfur. *International Interactions*, 1-24.

Pinto, C. M., Lopes, A. M., & Machado, J. T. (2012). A review of power laws in real life phenomena. *Communications in Nonlinear Science and Numerical Simulation*,*17*(9), 3558-3578.

Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. (2010). "Introducing ACLED-Armed Conflict Location and Event Data." Journal of Peace Research 47(5) 651-660.

Richardson, L. F. (1944). The distribution of wars in time. Journal of the Royal Statistical Society 57, 242– 250.

Richardson, L. F. (1948). Variation of the frequency of fatal quarrels with magnitude. Journal of the American Statistical Association 43, 523–546.

Richardson, L. F. (1960). Statistics of Deadly Quarrels. Pittsburgh: The Boxwood Press.

Roberts, D. C., & Turcotte, D. L. (1998). Fractality and self-organized criticality of wars.*Fractals*,*6*(04), 351-357.

Scharpf, A., Schneider, G., Nöh, A., & Clauset, A. (2014). Forecasting the risk of extreme massacres in Syria. *European Review of International Studies*,*1*(2), 50-68.

Schrodt Philip A (1991). Prediction of interstate conflict outcomes using a neural network. Social Science Computer Review 9(3): 359–380.

Schrodt Philip A, Davis Shannon G, Weddle Judith L (1994). Political science: KEDS – a program for the machine coding of event data. Social Science Computer Review 12(4): 561–587.

Schrodt, P. A., Yonamine, J., & Bagozzi, B. E. (2013). Data-based computational approaches to forecasting political violence. In Handbook of computational approaches to counterterrorism (pp. 129-162). Springer, New York, NY.

Shubik, M. (1997). Terrorism, technology, and the socioeconomics of death. *Comparative Strategy*,*16*(4), 399-414.

Simon, Herbert. A. 1955. On a class of skew distribution functions. Biometrika 42 (3/4): 425-40. United States Bureau of the Census. Census 2000. Washington, DC: Government Printing Office. http:// www.census.gov/main/www/cen2000.html

Singer, J. D. (1973). The peace researcher and foreign policy prediction. *The Papers of the Peace Science Society (International)*,*21*(1), 1-13.

Sipes, T., Jiang, S., Moore, K., Li, N., Karimabadi, H., & Barr, J. R. (2014). Anomaly Detection in Healthcare: Detecting Erroneous Treatment Plans in Time Series Radiotherapy Data. *International Journal of Semantic Computing*, *8*(03), 257-278.

Staniland, P. (2012). Between a Rock and a Hard Place: Insurgent Fratricide, Ethnic

Defection, and the Rise of Pro-State Paramilitaries. Journal of Conflict Resolution, 56 (1), 16–40.

Tomes, R. R. (2004). Relearning counterinsurgency warfare. *Parameters*, *34*(1), 16.

Trinn, C. (2015).*Konflikt und Komplexität: die Intensität innerstaatlicher Gewaltkonflikte in systemtheoretischer Perspektive*. Springer-Verlag.

Woodard, S. H., Bloch, G., Band, M. R., & Robinson, G. E. (2013). Social regulation of maternal traits in nest-founding bumble bee (Bombus terrestris) queens. *Journal of Experimental Biology*,*216*(18), 3474-3482.

Ziemke, J. (2012). Crisis mapping: The construction of a new interdisciplinary field?.Journal of Map & Geography Libraries,8(2), 101-117.

Zipf, George K. 1949. Human behavior and the principle of least effort. Cambridge, MA: AddisonWesley.

Abstract (Korean)

무력 충돌, 급진적인 테러 등을 포함하는 인간 폭등 (Human Insurgency)은 지금껏 만연하고 끊임없이 전세계적으로 위협을 주는 현상이다. 세계 개발 분야의 여러 주제 중 기관 및 조직의 주요한 관심은 이러한 충돌을 이해하고 모델 하는 것 이었다. 이러한 현상은 한 체제의 식량 안보, 아동 영양, 경제적, 그리고 환경적 문제와도 연관이 되는 것으로 간주되고 있다. 따라서 무력 충돌을 포함한 인간 폭등 현상을 매핑하는 것은 필수적으로 요구되고 있다.

기존의 연구들에서는 이런 현상들의 숨어 있는 패턴을 분석하고자 하는 시도들이 있었다. 그 중 두드러진 특징은 '멱법칙 (power law)' 분포를 따른다는 것이다. 멱법칙 분포는 꼬리가 긴 형태를 가지고 있다. 기존 정규분포와는 달리, 멱법칙 분포는 중심에서 멀리 위치하고 있는 현상이 더 이상 간소히 존재하는 것이 아니라, 상당히 존재한다고 볼 수 있다. 이러한 급증 규칙은 소수의 사건들이 작은 확률로 큰 규모로 발생하는 반면, 다수의 사건들은 작은 규모로 발생한다는 것을 암시하고 있다. 이러한 규칙들을 활용하여, 본 연구에서는 폭등을 매핑하기 위해 '비이상적 사건'들에 초점을 맞춘다. 비이상적 사건들은 멱법칙의 꼬리에 위치하는, 규모가 크지만 작은 확률로 발생하는 사건들을 일컫는다.

이러한 비이상적 사건들을 판별하기 위해, Variational Autoencoder (VAE) 라는

새로운 방법론을 도입하였다. 이 모델의 장점은 많은 량의 데이터를 처리하고, 비 선형적인 관계들을 포착할 수 있다는 것이다. 이러한 방법으로 data-driven 판별을 할 수 있게 한다. 이러한 모델로 훈련한 결과, False Negative (Type III error)와 False Positive (Type I error)이 발생하지 않았고, 비이상적인 사건들을 성공적으로 판별할 수 있었다. 본 연구에서는 이러한 모델을 제시함으로 실제 인도주의 단체나 정부에 적용되었을 경우, 제한된 자원을 효율적으로 배치하여 향후 폭등 규모를 완화할 수 있다고 본다.

주요어 : VAE, Machine Learning for development (ML4D), Anomaly Detection, Human insurgency, Armed conflict.
학 번 : 2017-28957