

# CHAPTER 1

## INTRODUCTION

### 1.1 Overview

Newspaper is an important part of our life as it is a printing media in which all information of either national or international news are published and delivered to the public every day. Henry Ward Beecher (1887), an American social reformer and well-known speaker once said, “Newspaper is a greater treasure to people than uncounted millions of gold.” Articles within the newspaper play essential role in education development and makes public aware about events happening in the region or nation that they are living in. By reading newspaper, readers can learn and observe others point of view as it brings another whole new perspective on same events. In the digital era, artificial intelligence (AI) expert make use of the digital technologies to get a better insight within the traditional media, newspaper.(Baker, 2019) Within these digital technologies, event prediction held a big portion as it forecast future events and it is valuable to alert public on predicted events.

Event prediction is a data analytic technique that make use of experience and knowledge as well as pattern from past to predict future events. Natural disaster prediction is one of the examples in applying concept of event prediction. For example, Japan is a well-known earthquake active country as its archipelago is in an area where multiple continental and oceanic plates collapsed together. Hence, earthquake forecasting is important for Japan and scientific report (D.Goltz, 2018) stated that earthquakes cluster in time and location as it can be predicted and take precaution before tragedy happened. Besides, event prediction is also useful in business intelligence. A report in 2016 also showed that business analysis and prediction help them to understand more about their customer, in order to enhance the success of their marketing strategies. (Erevelles *et al.*, 2016)

Every country had its obligation to protect its national security. National security is a requirement to maintain the survival of a country through economic, diplomacy, political and ethical power and focus on freedom from military threat and political coercion.. Without national security, a country might be at risk and attacks such as terrorism, sabotage, information warfare, etc might infiltrate the country. For example, ISIS threat stunned the world as a gunman, Mehdi Nemmouche opened fire at Jewish Museum of Belgium in Brussels as he is suspected in joining extremist groups, ISIS in Syria. This event took 4 lives of innocents.(Casert, 2016)

South China Sea (SCS) is a conflict zone whereby an estimated USD5 trillion worth of raw products shipped through shipping lanes in SCS each year and its nearby countries made them fight over each other to have the main control of the whole SCS.(?) The conflict is known as South China Sea disputes. The events of territorial disputes of South China Sea populated all the newspapers and many events regarding to the disputes were reported through national news agency. It rises concerns about the beginning of world war as for example a near-collision between US warship, Decatur and Chinese Luoyang missile destroyer in South China Sea highlights the escalating danger of confrontation between US and China. (Ni, 2018) SCS is strategically located at peripheral ocean that is a piece of the Pacific Ocean, starting from Karimata and Malacca Straits to the Strait of Taiwan with area about 3 500 000 km<sup>2</sup>. Besides, South China Sea is rich in marine life and natural resources such as oil and natural gas, even have the most of the important shipping lanes in the world. (E.Hayes, 1980)

Causality is the relationship between cause and effects. Every event will occur first on cause and followed by effects. In SCS disputes, causality is highlighted between benefits from SCS (cause) and territorial disputes (cause) is clearly highlighted in SCS disputes events.

In order to have an advanced insight among these disputes, event prediction is necessary, and causality should be taken as main attributes. However, there are still several problems and challenge to be solve in order to achieve an excellent prediction model based on causality.

## 1.2 Problem Background

National security is always the top priority of governments to protect society from disruption owing to a disaster or crisis. There are many aspects on national security such as territorial, economic, physical, social, political etc. However, the peaceful of national security had been affronted by South China Sea disputes.

Due to geological and resources advantages of South China Sea, countries within the region such as Brunei, China, Taiwan, Malaysia, Indonesia, Philippines, Vietnam etc. made competing the territorial claims over it. Based on news on The National Interest in 2016, an estimated US 5 trillion worth of global trade passes through the South China Sea annually. Hence, territorial disputes in the South China Sea started to concern worldwide community about peace of world. In order to claim the ownership of South China Sea, countries are challenging against each other by putting military force in the area. This can be observed from the news of China spent almost 1 year to build 7 new islands by moving sediment from the seafloor to reefs and after that focused on building ports, airstrips and other military structures on the islands.

South China Sea dispute had a brief background involving timeline from 221 BC until recent. Each of the historical event occurs and accumulates and eventually things go haywire. Many dispute events happened in either small or large scale. For example, Spratly island dispute (Gonzales, 2014) and “nine-dash” line (Zhen, 2014) that proposed by China is some of the significant disputes in South China Sea. Besides than these two issues, there are many issues that remain unsolved and will constantly concerning the worldwide community.

However, all the information retrieved from news article are unstructured. Unstructured data have no recognizable structure via pre-defined data models and schema and mainly generated by human or machine. (Taylor, 2018). By collecting these unstructured data from the past and analysing its trends, we are able to have a better understanding about what may happen in the future. (Bell, 2016). In SCS disputes, event prediction is important to give public a better understanding about future events that might happen. A better policy can be made with regards to protect national security

under SCS disputes with the event prediction technique based on unstructured data in news articles.

In event prediction based on news articles, there may have some challenges. For example, the event that causes another event might be completely different to each other.

There are several researchers research on topic event prediction with different method. Granroth-Wilding (2016) proposed a predictive neural network model that learns embeddings for words describing events, a function to change embeddings into event representation and a function to predict the degree of relationship between two events. However, the model is more focus on chain or events sequence which is good for rich-informative events but might not suitable for news articles that have unordered sequence. Preethi (2015) proposed an event prediction model for Tweets using temporal sentiment analysis and causal rules extraction. This model is useful to analyse user's sentiments and predict future events using temporal attribute. This study analyse sentiments of user's opinion and is not suitable for news articles whereby formal news reports seldom express their sentiments within the articles.

Radinsky (2014) proposed a predictive model based on causality attribute. The predictive model is suitable for unstructured data such as news article and learn from previous events to predict the future. For example, Given a South China Sea events stated that "China (Actors) construct (Actions) a new military purpose platform (Objects) on strategically-located Bombay Reef, South China Sea (Location) in 22 Nov 2018 (Time)". From this event, we can observe clearly that there are 5 tuples <Actors, Actions, Objects, Location, Time>. From the 5 tuples, we can extract useful keywords that identify the characteristics of the events. The further event happened after this is that "Vietnam's sovereignty in South China Sea is threatened and Vietnam started to protest". This causality pair indicated the relationship between cause "Country expand their territorial area in South China Sea" and effect "Another country's sovereignty is threatened and started to protest". This model makes use of previous events to predict future events based on causality attributes.

### **1.3 Problem Statement**

Online news is unstructured data and it is a challenge to extract correct information from online news articles that contain different resources. Event prediction is also biased to measure, and it is a great challenge to achieve a good predictor.

### **1.4 Aim/Purpose of Study**

South China Sea (SCS) is a conflict zone where events happened from time to time with different severity. This greatly impact or influence the policy that made by government of Malaysia to overcome the negative impacts brought by SCS territorial disputes in term of national security. This study will address the SCS disputes issues by implementing Radinsky (2012) predictive model using Pundit algorithm. This predictive model learn causality of events and provide accurate suggestion based on predicted outcome.

### **1.5 Objective**

1. To extract keywords from news articles using 5 tuples approach on <Actors, Actions, Objects, Location, Time> for better event representation.
2. To produce generalized node and path from extracted keyword and finding minimal generalization path for generalizing extracted information.
3. To produce abstraction tree (AT) based on events nodes and causality prediction rules in order to building a suitable prediction model and evaluate the accuracy of prediction model with precision and recall or comparing with human predictor.

### **1.6 Scope**

Event prediction is important to have better understanding for future event that may happen, but it also have its limitations that need to address and narrow down. Below are the scope of this study:-

1. The coverage of this study will be using news articles that extracted related to SCS conflict from the national news agency.

2. The study will focus on SCS conflict news from Vietnam News Agency (VNA) and Xinhua News Agency (China)
3. Prediction model in this study will be based on Pundit Algorithm proposed by Radinsky (2012).
4. The study will only covers the causality prediction rules that adopted from Radinsky (2012).
5. 5 tuples approach will be used to extract the keyword on <Actors, Actions, Objects, Location, Time>

### **1.7 Significant of Study**

Event prediction is important because it give a better insight and forecast for public about events that possible to happen. Besides, in SCS conflict, event prediction can be used to protect national security of Malaysia and appropriate actions can be took by governments to prevent the happening of unseen tragedy based on predicted events. In addition, Malaysia governments are able to protect the sovereignty of country in SCS region to claim peace and avoid involving in SCS conflict.

### **1.8 Organisation of Study**

**Chapter 2** will be discussed about the literature review of event prediction and information extraction from news articles based on different method and referenced method. **Chapter 3** will focus on research methodology, research framework, overall research phase, and measurement and rules. **Chapter 4** will be present about initial result of implementing prediction model proposed by Radinsky (2012) by on step-by-step approach on causality attributes.

## **CHAPTER 2**

### **LITERATURE REVIEW**





## REFERENCES

- Baker, J. (2019). *Machine Learning Versus The News*. Retrievable at <https://towardsdatascience.com/machine-learning-versus-the-news-3b5b479d8e6a>.
- Bell, B. (2016). *The Road to Prediction: Using Unstructured Data*. Retrievable at <https://www.dataversity.net/the-road-to-prediction-using-unstructured-data/>.
- Casert, R. (2016). *Belgium ramps up security for lone suspect in Jewish Museum attack*. Retrievable at <https://archive.is/20140527190719/http://www.theglobeandmail.com/news/world/police-hunt-brussels-jewish-museum-shooter-france-tightens-security/article18835439/#selection-583.1-583.67>.
- D.Goltz, J. (2018). *Earthquake advisories can save lives*. Retrievable at <https://www.japantimes.co.jp/opinion/2018/04/17/commentary/japan-commentary/earthquake-advisories-can-save-lives/#.XIc8d7NbzeQ>.
- E.Hayes, D. (1980). South China Sea, PACIFIC OCEAN. *The Tectonic and Geologic Evolution of Southeast Asian Seas and Islands*.
- Erevelles, S., Fukawa, N. and Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*. 69(2), 897–904.
- Gonzales, R. (2014). The Spratly Islands Dispute: International Law, Conflicting Claims, and Alternative Frameworks For Dispute Resolution.
- Taylor, C. (2018). *Structured vs. Unstructured Data*. Retrievable at <https://www.datamation.com/big-data/structured-vs-unstructured-data.html>.
- Zhen, L. (2014). What's China's 'Nine-Dash Line' and why has it created so much tension in the South China Sea? *South China Morning Post*.