Application of Knowledge-oriented Convolutional Neural Network For Causal Relation Extraction In South China Sea Conflict Issues

Koh Leong Chien
Cyber Threat Intelligence Lab,
Information Assurance & Security Research Group (IASRG)
School of Computing, Faculty of Engineering
Universiti Teknologi Malaysia
81310 Johor Bahru, Johor, Malaysia
lckoh0830@gmail.com

Anazida Zainal
Cyber Threat Intelligence Lab,
Information Assurance & Security Research Group (IASRG)
School of Computing, Faculty of Engineering
Universiti Teknologi Malaysia
81310 Johor Bahru, Johor, Malaysia
anazida@utm.my

Fuad A. Ghaleb
Cyber Threat Intelligence Lab,
Information Assurance & Security Research Group (IASRG)
School of Computing, Faculty of Engineering
Universiti Teknologi Malaysia
81310 Johor Bahru, Johor, Malaysia
abdulgaleel@utm.my

Mohd Nizam Kassim
Cybersecurity Malaysia
Level 7 Tower 1, Menara Cyber Axis, Jalan Impact,
63000 Cyberjaya, Selangor Darul Ehsan, Malaysia
nizam@cybersecurity.my

2021 3rd International Cyber Resilience Conference (CRC) | 978-1-6654-1844-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/CRC50527.2021.9392525

Abstract*—Online news articles are an important source of information for decisions makers to understand the causal relation of events that happened. However, understanding the causality of an event or between events by traditional machine learning-based techniques from natural language text is a challenging task due to the complexity of the language to be comprehended by the machines. In this study, the Knowledge-oriented convolutional neural network (K-CNN) technique is used to extract the causal relation from online news articles related to the South China Sea (SCS) dispute. The proposed K-CNN model contains a Knowledge-oriented channel that can capture the causal phrases of causal relationships. A Data-oriented channel that captures the position information was added to the K-CNN model in this phase. The online news articles were collected from the national news agency and then the sentences which contain relation such as causal, message-topic, and product-producer were extracted. Then, the extracted sentences were annotated and converted into lower form and base form followed by transformed into the vector by looking up the word embedding table. A word filter that contains causal keywords was generated and a K-CNN model was developed, trained, and tested using the collected data. Finally, different architectures of the K-CNN model were compared to find out the most suitable architecture for this study. From the study, it was found out that the most suitable architecture was the K-CNN model with a Knowledge-oriented channel and a Data-oriented channel with average pooling. This shows that the linguistic clues and the position features can improve the performance in extracting the causal relation from the SCS online news articles. Keywords-component; Convolutional Neural Network, Causal Relation Extraction, South China Sea.

***Keywords*—Knowledge-oriented Convolutional Neural Network, South China Sea Conflict, Causal Relation Extraction**

I. Introduction

Causality is the relationship between cause and effect. It also plays an important role in explanation and decision making [1] in which every decision made may have different effects. To protect national security, one of the methods is to learn the causality from the previous events and implement preventive measures. This is because the causal relation allows people to understand the pattern of past events in which different effects may be caused by the same cause. From that, the pattern of an event and the correlation between the event can be learned. Although some event correlations may not necessarily contain causality, causation still implies correlation [2]. Therefore, learning correlation will help people to understand the complex relationship between event and causal relation to make a prediction.

The focus of this paper is to learn the casual relation from online news articles more specifically is the articles related to the South China Sea (SCS) conflict. This is because it is the world busiest waterways subject to several overlapping territorial disputes involving several countries such as China, Philippines, Vietnam, Indonesia, Malaysia, Taiwan, and Brunei due to both islands and maritime claims among these countries. This is because SCS is a key commercial as one-third of the world's shipping is passaging through it. The conflict recently emerged due to the oil discovery beneath the South China Seawater. Learning the causal relation from the SCS event is essential to help the relevant countries in learning the cause and effect in order to protect national as the issues that happened in SCS had threatened national security. A lot of information can be obtained easily by using online sources such as online newspapers or articles. National News Agency of different countries such as Malaysia and China have their own websites to update the news including their stand. From there, the stand of other countries in SCS can be known and they can be the references in making decisions to secure the country from the SCS dispute.

A. Problem Background

Events are not happening suddenly, there must be a pattern or reason to lead an event to occur. Learning causal relations from an event is significant to understand or predict the result of an action or a random occurrence in which the causality may give an impact on reasoning and decision-making for humans [3] [4]. The implementation of preventive measures is essential to prevent the previous event happened again. Causal relation extraction is a technique to extract the cause-effect relation from the sentence. This technique is used to help people to understand the causal relation from the huge amount of unstructured online news articles easily.

In 2012, Ackerman [5] extracted a causal network of news topics, and the Palladian toolkit was used to extract the information from the online news article using a web search engine such as Lucene and used Reuters corpus as a corpus. He was using news topic references (NTR) to find the related news on Wikipedia. In his work, the news topic was first given to collect the multiple or similar news articles through a search engine and then the seed was extended with causal markers to find the articles that contain causal terms such as "causal" or "led to". The collected articles were then split into sentences and processed to be presented in the causal graph. However, many manual works were conducted before extraction which leads to low extraction accuracy due to the imprecise manual work. In 2012, Ackerman [5] extracted a causal network of news topics, and the Palladian toolkit was used to extract the information from the online news article using a web search engine such as Lucene and used Reuters corpus as a corpus. He was using news topic references (NTR) to find the related news on Wikipedia. In his work, the news topic was first given to collect the multiple or similar news articles through a search engine and then the seed was extended with causal markers to find the articles that contain causal terms such as "causal" or "led to". The collected articles were then split into sentences and processed to be presented in the causal graph. However, many manual works were conducted before extraction which leads to low extraction accuracy due to the imprecise manual work.

In 2013, Sorgente [6] had proposed automatic extraction of extraction relation from natural language text while Yang and Mao [7] proposed multi-level causal relation identification in which any verb and preposition based on the linguistic structures was used to extract the causal relation. In 2016, Zhao [8] extracted event causality by extra features such as position, syntactic, and connective to improve the performance. After 2016, the deep learning approach was proposed due to a large amount of data is needed in extraction. In 2017, Silva [9] used Convolutional Neural Network (CNN) to extract causal relations while in 2018, Dasgupta [10] proposed automatic extraction of causal relations using a deep neural network. In 2019, Li and Mao [11] improved their previous work by using a Knowledge-oriented convolutional neural network (K-CNN) that contains two convolutional channels to capture more accuracy in syntactic and semantic analysis of natural language text.

Nowadays, an abundant amount of online news articles is available and can be obtained easily but they are unstructured and may conflict. Thus, it may take a lot of time to understand the causality from them due to the noise and the redundancy. Besides that, SCS news has involved different events since many years ago. Therefore, plenty of time is needed to learn the causal relation from the SCS news.

To this end, in this study, the K-CNN that was proposed by Li and Mao [11] is modified to extract causal relations from the SCS online news articles. Both data-oriented channel and Knowledge-oriented have been modified by experimenting with different forms of pooling layers in the deep learning architecture. Several experiments have been conducted to evaluate the effectiveness of the modified K-CNN architecture Besides, the impact of excluding the data-oriented channel on the model effectiveness has been also studied.

B. Problem Statement and aim

Online news articles are unstructured and it is challenging to understand the causal relationship of an event. There are some existing works for causal relation extraction such as the rule-based method, machine-learning-based method, or deep learning method. The rule-based method has low accuracy because it cannot deal with the complexity of causal relations in natural language. There are many rules and pattern the matching needed which are massive works. For the machine learning method, the performance of the proposed works is relying on external NLP toolkits such as named entity recognizer, POS tagger, and dependency parser. Deep learning is promising in extracting the causal relations from text data; however, it has not yet received deep investigations and research. Therefore, this research aims at improving the performance of the K-CNN that was proposed by Li and Mao [11] to extract causal relations from the SCS online news articles.

II. Preliminaries And Related Work

A. Relation Extraction

Relation Extraction is a subfield of Information Extraction (IE) and it is used to extract the relationship from the natural language text such as causality and relationship between human and object (Li and Mao, 2019). In order to extract the relation, a relational three tuples <Entity1, Relation, Entity2> was used, in which Entity1 and Entity 2 are entity types while Relation is relation description (Zhang et al., 2017). For example, "Mark Zuckerberg is the founder of Facebook", the relationship can be extracted between Facebook and Mark Zuckerberg as the founder. Relation extraction can be used in automatic question answering system to answer the question (Zhang et al., 2017) as in the rapid development of information technology, people may have difficulty in obtaining useful information from the huge volume of unstructured online data such as raw-text from email, report, resumes, papers or online news. This may help people in understanding the relationship between two entities easily and increase the efficiency when training data as it can extract the relationship from many text or natural language texts with the same pattern.

B. South China Sea

In the South China Sea, there are a lot of resources such as oil and natural gas which are high in value and this becomes the main cause of many countries such as China, Philippines, Vietnam, and Malaysia claimed that the area is their territory. However, every country's claim is not approved by other countries. Many reclamations of land or island happened in SCS dispute areas such as China and Vietnam overclaimed that the Paracel island in the SCS area. Many incidents happened in the overlapping claims to the SCS area. For example, in March 2012, China had detained 21 Vietnamese fishermen as they are fishing around the disputed area illegally, and China is requesting USD 11,000 in order to release them home (Kotte, 2012). However, Vietnam refused to fulfill China's request because Vietnam claimed that they were captured in their own territories and China should release them unconditionally. Moreover, in 2014 there is a malware called "South China Sea Remote Access Trojan (RAT)" or NanHaiShu which was believed to involve in the

SCS dispute was discovered spied on the Philippines and other organizations involved in assisting the South East Asian country [12]. This malware was spread by spear-phishing e-mail messages that contain the malware as a file attachment and it can collect the sensitive data from the infected PC to the target server [13]. According to the F-Secure's report, they believed that the threat actor of the malware is China after performing a technical malware analysis on the NanHaiShu in which the data was sent to a remote command and control server with a Chinese IP address [14]. These issues had jeopardized the national security of the victim country as there was the involvement of the death of people and the loss of the credential information.

C. Causal Relation Extraction

Causal Relation Extraction is a method to detect and extract the cause-effect relationship automatically from the natural language text. It is a very challenging task in Artificial Intelligence because there is a lot of different cue phrase and the complexity of causal relation expression from the natural language text. The model or machine needs to deal with the interplay of syntax, continually evolving vocabulary, semantics, and ambiguous constructs such as metaphors, sarcasm, and slang [11]. The causal relation extraction is widely used in many fields and it plays significant roles in applications such as information retrieval, question answering, event reasoning, and predictions [11] that can identify the relation between events by constructing the causal chains [16]. Figure 1 shows the general process of causal relation extraction. There are many current methods to extract the causal relation which are 1) Rule-based, 2) Machine-learning and 3) Deep learning method.

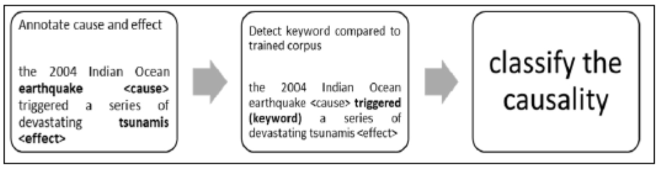


Fig. 1 General Process of causal relation extraction

D. Rule Based Method

In the 80s, there is already have the idea of extracting information from natural language texts automatically [16]. Selfridge [17] stated that the knowledge representation model with a semantic framework is a very important requirement to build an effective causal classifier, and to identify the causal relation from the text accurately, it must have the domain knowledge simultaneously. Moreover, the model needs to be able to recognize the words or components between the cause and effect to construct a comprehensive causal chain with high accuracy. In the new millennium, the works focus on the large and domain-independent texts, automation, and scalability as previous works consist of many manual works such as constructs the linguistic patterns and encoded propositions. Girju [18] proposed a semi-supervised method to discover the linguistic patterns expressing causal relations automatically to validate the extracted patterns based on some constraints on NP1, NP2, and causal-verb. First, they generate syntactic patterns in the form of <NP1 causal-verb NP2> by searching a collection of texts on the Internet which is from lexical knowledge base WordNet [19] where Wordnet is a lexical database for the English language. After that, the extracted patterns were validated and ranked on a scale of one to four, with one is highly likely causal relation and 4 is high ambiguity through a WordNet-based course-grained process. This work can replace the manual works and involved large data however there are some limitations also which is they focused on extracting explicit and simple causative verb forms only.

E. Machine Learning Method

As the number of data increases, to extract implicit patterns in text automatically, the machine learning method is more suitable than the rule-based method. Recently, Sorgente [6] extracted the possible cause-effect pairs from a sentence by matching the pattern based on simple causative verbs, phrasal verbs, noun+preposition, passive causative verbs, and special single prepositions. Then, a Bayesian classifier and Laplace smoothing was used to discard the correct pairs. Yang and Mao [7] proposed a multilevel relations extractor (MRLE) which is based on the linguistic knowledge of the dependency grammar and constituent grammar that can extract all potential causal relations. WordNet, Verbnet, and FrameNet had been used as the lexical knowledge bases and classification is done by using restricting boosting and SVM with the RBF kernel. Zhao [8] proposed a Restricted Hidden Naïve Bayes (RHNB) model which is an improvement from Hidden Naive Bayes (HNB) [20] to extract the causal relations among the features based on causal connectives. They divided the features into four groups which are contextual, syntactic, position, and connective but have the same hidden parent.

F. Deep Learning Method

Rule-based and traditional machine learning techniques have their limitations once they are being used for extracting the causal relations from unstructured and human language data. Deep learning had been implemented in relation extraction that can minimize the reliance on NLP for feature acquisition [11]. Deep learning is using neural network architecture which contains many simple and connected neurons, but it has more hidden layers of neural networks than the traditional neural network. In the beginning, each of the neuron as an input unit will receive vary of input data in order to recognize and analyses them. At the end of the neural network, there is an output unit to represent how it responds to the information that has been learned. In between of input unit and output unit, there is one or more hidden layer which acts like the human brain to process the input data based on specific function according to the input data type. With the more hidden neural network, deep learning improves the classification performance as it takes a huge number of datasets to train a model which able to learn more features from the data. Deep learning is widely used in image processing such as automated driving and also text classification that brings a lot of benefits and convenience to humans. One of the famous types of deep learning is Convolutional Neural Network (CNN).

G. Convolutional Neural Network (CNN)

CNN is a feed-forward neural network which performs convolutional and pooling process and the neural network is organized in three-dimensional layers which are different from the traditional neural network. CNN is widely used in image processing previously for better performance in face recognition and automated driving. Recently, Zeng [21] implement the position embedding technique in relation classification which determines the relative distance between two words, if the relative distance is too far, hence the sentence may contain relations. de Silva [9] used the CNN technique and knowledge-based feature to identify the causality from the natural language text. Li and Mao [11] proposed a Knowledge-oriented convolutional neural network (K-CNN) which contains two channels in K-CNN: a Knowledge-oriented channel and a data-oriented channel. They improve the performance of CNN in causal relation extraction which they adopt Zeng's [21] position embedding to find the relative distance between cause and effect and reduce dimensionality by remove irrelevant words and cluster similar words.

III. METHODOLOGY

Figure 2 shows the research framework in this study. There were three phases which were 1) Data Pre-Processing, 2) Causal Relation Extraction, and 3) Comparison with different Architecture of K-CNN model.

A. Dataset

The dataset focused on the online news articles related to the SCS issues from China, Philippines, and Vietnam. The news was collected from the national news agency such as *China Daily News* that represents the government's stand. Then, the sentences that contain relation such as "Cause-Effect", "Message-Topic" and "Product-Producer" were extracted and annotated to train the model. These relationships were chosen by referring to the relationship labelled in the SemEval 2010 Task8 dataset [12]. Table 1 shows the description of the relationship used in this study. This study focused on causal relation extraction only, therefore other relations were labeled as "Other". Figure 3 shows the sample of the online news articles collected and Figure 4 shows the sample of causal phrases extracted and annotated. Table 1 shows the data distribution of data. A total of 192 sentences were collected from the online news which contains three relationships as shown in Table 2.

B. Performance Measure

In this study, the performance measure was done by using the precision, recall, and F1-score of the model. The formula as shown below:

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{1}$$

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{2}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{3}$$

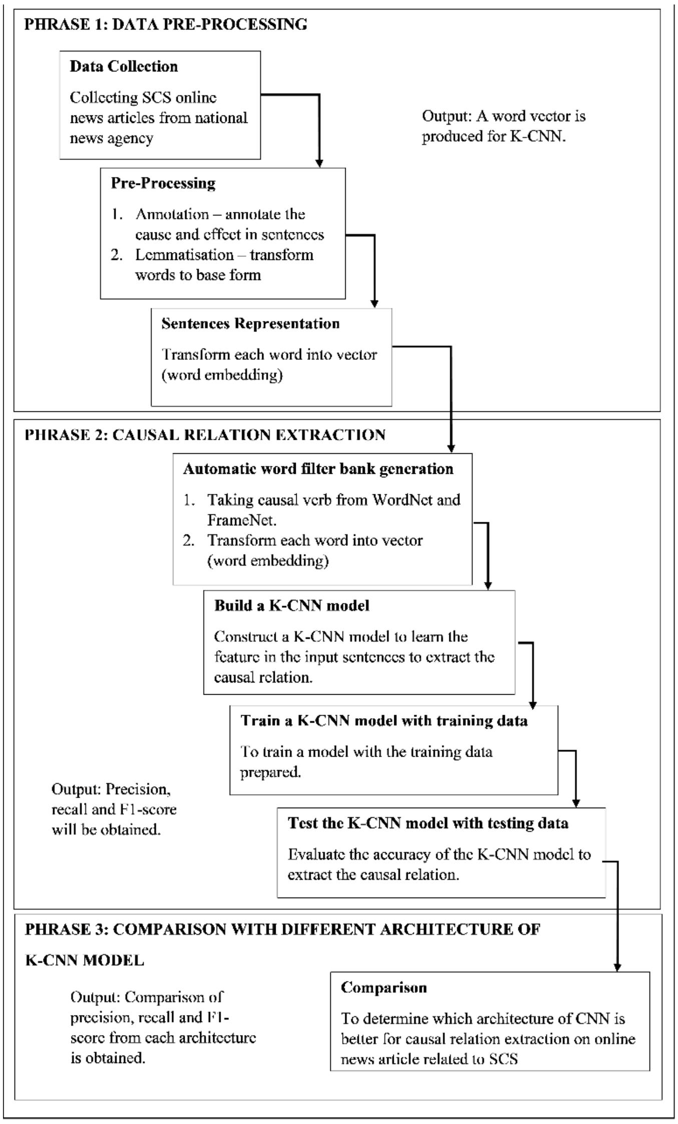
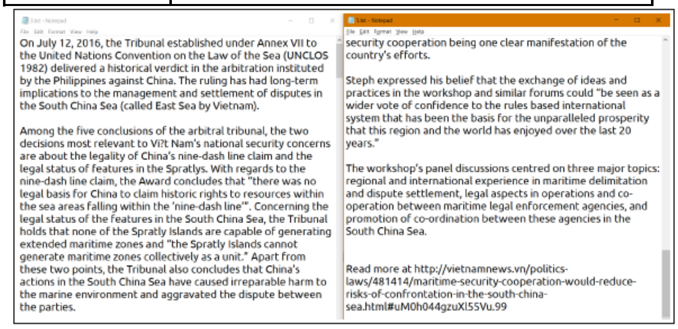


Fig. 2: Research Framework

TABLE 1: RELATION DESCRIPTION

Cause-Effect	The cause is followed by effect in the sentences.
Effect-Cause	The effect is followed by cause in the sentences.
Other	The non-causal sentences or no causal relation sentences.



On July 12, 2016, the Tribunal established under Annex VII to the United Nations Convention on the Law of the Sea (UNCLOS 1982) delivered a historical verdict in the arbitration instituted by the Philippines against China. The ruling has had long-term implications to the management and settlement of disputes in the South China Sea (called East Sea by Vietnam).

Among the five conclusions of the arbitral tribunal, the two decisions most relevant to Việt Nam's national security concerns are about the legality of China's nine-dash line claim and the legal status of features in the Spratlys. With regards to the nine-dash line claim, the Award concludes that "there was no legal basis for China to claim historic rights to resources within the sea areas falling within the 'nine-dash line'". Concerning the legal status of the features in the South China Sea, the Tribunal holds that none of the Spratly Islands are capable of generating extended maritime zones and "the Spratly Islands cannot generate maritime zones collectively as a unit." Apart from these two points, the Tribunal also concludes that China's actions in the South China Sea have caused irreparable harm to the marine environment and aggravated the dispute between the parties.

security cooperation being one clear manifestation of the country's efforts.

Steph expressed his belief that the exchange of ideas and practices in the workshop and similar forums could "be seen as a wider vote of confidence to the rules based international system that has been the basis for the unparalleled prosperity that this region and the world has enjoyed over the last 20 years."

The workshop's panel discussions centred on three major topics: regional and international experience in maritime delimitation and dispute settlement, legal aspects in operations and co-operation between maritime legal enforcement agencies, and promotion of co-ordination between these agencies in the South China Sea.

Read more at http://vietnamnews.vn/politics-laws/481414/maritime-security-cooperation-would-reduce-risks-of-confrontation-in-the-south-china-sea.html#uM0h044gzuXl55Vu.99

Fig. 3 Sample of online news articles collected

Fig. 4 Sample of causal phrases extracted and annotated

TABLE 2: DATA DISTRIBUTION

	Cause-Effect	Effect-Cause	Other	Total
Training data	75	54	23	**152**
Testing data	22	7	10	**39**
Total	85	61	33	**192**

IV. RESEARCH DESIGN AND IMPLEMENTATION

The proposed solution was developed in three phases where each phase corresponds with the objectives of this study. Figure 5 shows the overflow of the experiment.

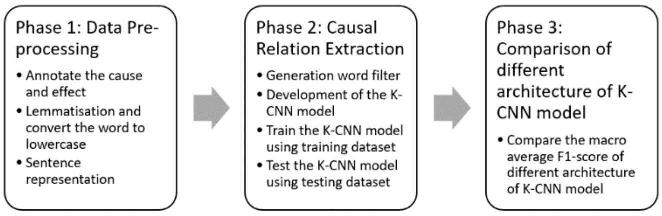


Figure 5 overflow of the experiment

A. Data Pre-processing

As discussed in the Methodology, the sentences extracted were annotated with two entities and labeled with their relation by using Brat Annotation tools. Then all the words were transformed into lower case and base form by using WordNet Lemmatizer. Next, each word was converted to a vector by looking up the word embedding table. The pre-trained word embedding table using English Wikipedia which was taken from https://www.cs.york.ac.uk/nlp/extvec/ [22] was used as the word embedding table in this study.

B. Causal Relation Extraction

This phase consists of two stages as follows:

1. Word Filter: In this stage, a word filter was generated by collecting the causal phrases from the FrameNet. The FrameNet categorizes English words and sentences into different frames by describing a type of event, relation or object. 39 causal frames from the FrameNet had been found such as "Causation", "Causation_scenario", "Triggering", "Reason", "Explaining_the_facts", "Response" and other 33 frames that start with "Cause", and their lexical units were collected to generate the word filter. This word filter was used as a convolutional filter in the Knowledge-oriented channel in the next stage.

2. Development of K-CNN model: In this stage, the K-CNN model was built by using PyTorch which is a python deep-learning framework built on top of Tensor. Then the layers in the CNN model were defined custom which contains the Embedding layer, Convolutional layer, Pooling layer and Output layer. In Embedding layer, the word vector was used to convert the word into the vector by looking at its index. Then in the Convolutional layer, named as knowledge-oriented channel produced the features map of the dataset. The Knowledge-oriented channel took the words between the two entities as input and then calculated the cosine similarity between the word filters and the input. Equation (4) shows the equation used in the Convolution layer.

$$m_i = \left(\sum_{j=1}^{k} f_j^T \; W_{i+j-1} + b \right) / k \tag{4}$$

After that, it extracted the most significant features in the Pooling layer by using Max-pooling. The final feature vector produced was passed to a classifier to classify the relation. After K-CNN was built, it was trained and tested with the training and testing data.

C. Comparison of different architecture of K-CNN model

To find out the most suitable architecture of the K-CNN model for the causal relation extraction from SCS online news article, the different architecture of the K-CNN model proposed was tested. The Max-pooling function in the K-CNN model that was built in the previous stage was changed into average-pooling to compare the performance. Then, a data-oriented channel was added to the K-CNN model. This channel was used to extract the position features from the sentences which is the distance between the entities and other words. The formula used for this channel is shown in (5).

$$m_i = tanh\left(\sum_{j=1}^{k} f_j^T \; W_{i+j-1} + b \right) \tag{5}$$

This channel took the whole sentences as input data which is different from the Knowledge-oriented channel. Next, the pooling function in both channels was exchanged with average-pooling also to compare the performance.

V. RESULT, ANALYSIS AND DISCUSSION

A. K-CNN model (base model)

The K-CNN model that contains only the Knowledge-oriented channel and using the Max-pooling function was the base model. Table 3 shows the result obtained from the K-CNN model.

TABLE 3: K-CNN MODEL'S RESULT

Class	Precision (%)	Recall (%)	F1-score (%)
Cause-Effect	60.00	90.00	72.00
Effect-Cause	60.00	75.00	66.67
Macro-average F1-score	60.00	82.50	69.33

The performance for class Cause-Effect is good because the Knowledge-oriented channel can capture the causal phrases such as "cause", "lead to" and "result in". However, for the class Effect-Cause, the performance is low because there are some sentences that contain the "because" but do not contain the causal relation. For example, "*A strong navy and a strong air force are crucial to safeguarding China's sovereignty in the South China Sea" because a mighty military is a deterrence, Zhou Yongsheng, a professor of international relations at China Foreign Affairs University, told the audience at the ceremony on Thursday.*", this sentence contains the keyword "because" but the sentence is not causal, it is more to elaborate the point.

B. Comparison with different architecture of K-CNN model

For a better explanation, each of the architecture was labelled with a name as shown below:

1) K_CNN_Kmax: base K-CNN model with Knowledge-oriented channel only and using Max-Pooling layer.
2) K_CNN_Kavg: K-CNN model with Knowledge-oriented channel only and using Average-Pooling layer.
3) K_CNN_Kmax_Dmax: K-CNN model with Knowledge-oriented and Data-oriented channel (Both channel using Max-pooling). This architecture was proposed by Li and Mao's work [11] which will be used as a baseline for validating the other architectures that were proposed by the authors in this study.
4) K_CNN_Kavg_Dmax: K-CNN model with Knowledge-oriented (using Average-Pooling) and Data-oriented channel (using Max-Pooling).
5) K_CNN_Kmax_Davg: K-CNN model with Knowledge-oriented (using Max-Pooling) and Data-oriented channel (using Average- Pooling).
6) K_CNN_Kavg_Davg: K-CNN model with Knowledge-oriented and Data-oriented channel (Both channel using Average-pooling).

TABLE 4: SUMMARY OF RESULTS OF THE DIFFERENT ARCHITECTURE OF K-CNN MODEL WITH F1-SCORE AND MACRO-AVERAGE F1-SCORE

Architecture	F1-Score (%)		Macro-Average F1-score (%)
	Cause-Effect	Effect-Cause	
K_CNN_Kmax	72.00	66.67	69.33
K_CNN_Kavg	76.19	72.72	74.46
K_CNN_Kmax_Dmax	78.26	72.72	75.49
K_CNN_Kavg_Dmax	**85.71**	72.72	79.22
K_CNN_Kmax_Davg	76.19	72.72	74.46
K_CNN_Kavg_Davg	81.82	**88.89**	**85.35**

From Table 4, the K_CNN_Kavg has better performance than K_CNN_Kmax. This is because some of the causal sentences in the news articles do not contain causal keywords. Therefore, average pooling is better for the Knowledge-oriented channel as it can keep the information of every word while max-pooling took the highest value as the feature only. Besides that, the K_CNN_Kmax_Dmax model shows that it performs slightly better than the base model which means that the Data-oriented channel can help to increase the performance of the base model. This is because there is some of the causal keywords do not lie in between the two entities such as "*As a result*" and "*Due to*". However, the K_CNN_Kavg_Dmax model increase much performance compare to the base model and achieve the highest F1-score for the class "Cause-Effect". This confirms that the Knowledge-oriented channel with an average-pooling function can work well with the Data-oriented channel. From the last two models (as listed in Table 4), the K_CNN_Kmax_Davg and the K_CNN_Kavg_Davg, they showed that the Data-oriented channel with an average-pooling increase the performance well where K_CNN_Kavg_Davg achieved the highest score for F1-score of "Effect-Cause" and macro-average F1-score. K_CNN_Kmax_Davg also showed that the Knowledge-oriented channel with max-pooling cannot work well with Data-oriented channel as Data-oriented channel with average-pooling increases the performance but max-pooling in Knowledge-oriented channel dropped the performance. In short, the most suitable architecture for this study was the K_CNN_Kavg_Davg model.

VI. CONCLUSION

In conclusion, this study has studied the application of K-CNN to extract the causal relation extraction from the SCS online news articles by referring to Li and Mao's work [11]. This can help people or the government to understand easily the causal events that happened in the SCS disputes to increase awareness of national security. In this study, the K-CNN that was proposed by Li and Mao [11] is modified to extract causal relations from the SCS online news articles. Both data-oriented channel and Knowledge-oriented have been modified by experimenting with different forms of pooling layers in the deep learning architecture. Several experiments have been conducted to evaluate the effectiveness of the modified K-CNN architecture Besides, the impact of excluding the data-oriented channel on the model effectiveness has been also studied. In the first phase, the causal sentences had been annotated with two entities and labeled with their relation. Then, every word was transformed into lowercase and base form and eventually converted into a vector by looking up the pre-trained word embedding table. In phase two, a word filter was built by collecting the lexical units from the frame that related to causal in the FrameNet. Then, the K-CNN model was constructed, trained, and tested with the training and testing data. Finally, six different architecture of the K-CNN model were compared with their performance by comparing the macro-average F1-score. The most suitable architecture model was found out throughout this research which performs well in the extraction of causal relation from online news articles is the K_CNN_Kavg_Davg model.

VII. ACKNOWLEDGMENT

We would like to thank Universiti Teknologi Malaysia and UTM RMC for funding this project under UTM TDR Grant PY/2018/03545.

REFERENCES

[1] Glymour, C., Spirtes, P., Zhang, K. and Schölkopf, B. (2017). Learning causality and causality-related learning: some recent progress. National Science Review. 5(1), 26–29. ISSN 2095-5138. doi:10.1093/nsr/nwx137.

[2] Stanovich, K. E. (1992). How to think straight about psychology. HarperCollins

[3] Goldvarg, E. and Johnson-Laird, P. (2001). Naive causality: A mental model theory of causal meaning and reasoning. Cognitive Science. 25, 565–610. doi:10.1016/ S0364-0213(01)00046-5.

[4] Peña, A., Sossa, H. and Gutiérrez, A. (2008). Causal knowledge and reasoning by cognitive maps: Pursuing a holistic approach. Expert Systems with Applications. 35(1-2), 2–18.

[5] Ackerman, E. J. M. (2012). Extracting a causal network of news topics, 33–42

[6] Sorgente, A., Vettigli, G. and Mele, F. (2013). extraction of cause-effect relations in natural language text. 1109, 37–48.

[7] Yang, X. and Mao, K. (2014). Multi level causal relation identification using extended features. Expert Systems with Applications. 41(16), 7171 – 7181. ISSN 0957- 4174. doi:https://doi.org/10.1016/j.eswa.2014.05.044. Retrievable at http://www. sciencedirect.com/science/article/pii/S0957417414003303.

[8] Zhao, S., Liu, T., Zhao, S., Chen, Y. and Nie, J.-Y. (2016). Event causality extraction based on connectives analysis. Neurocomputing.

173, 1943 – 1950. ISSN 0925- 2312. doi:https://doi.org/10.1016/j.neucom.2015.09.066. Retrievable at http:// www.sciencedirect.com/science/article/pii/S0925231215013995.

[9] de Silva, T. N., Zhibo, X., Rui, Z. and Kezhi, M. (2017). Causal Relation Identification Using Convolutional Neural Networks and Knowledge Based Features. World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering. 11(6), 697–702.

[10] Dasgupta, T., Saha, R., Dey, L. and Naskar, A. (2018). Automatic Extraction of Causal Relations from Text using Linguistically Informed Deep Neural Networks. In Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue. 306–316.

[11] Li, P. and Mao, K. (2019). Knowledge-oriented convolutional neural network for causal relation extraction from natural language texts. Expert Systems with Applications. 115, 512–523.

[12] Bureau, O. (2016). Malware Called 'South China Sea RAT' Spied on Philippines, Other Targets. Retrievable at http://www.defenseworld.net/news/16792/Malware_Called__South_China_Sea_RAT__Spied_on_Philippines__Other_Targets.

[13] Cimpanu, C. (2016). Chinese APT Deploys NanHaiShu RAT Against International Adversaries. https://news.softpedia.com/news/chinese-apt-deploys-nanhaishu-ratagainst- international-adversaries-506984.shtml.

[14] F-secure (2016). NANHAISHU - RATing the South China Sea whitepaper. F-Secure.

[15] Hendrickx, I., Kim, S. N., Kozareva, Z., Nakov, P., Séaghdha, D. O., Padó, S., Pennacchiotti, M., Romano, L. and Szpakowicz, S. (2019). Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. arXivpreprint arXiv:1911.10422.

[16] Asghar, N. (2016a). Automatic Extraction of Causal Relations from Natural Language Texts: A Comprehensive Survey. CoRR. abs/1605.07895. Retrievable at http://arxiv.org/abs/1605.07895.

[17] Selfridge, M. (1989). Toward a natural language-based causal model acquisition system. Applied Artificial Intelligence an International Journal. 3(2-3), 191–212.

[18] Girju, R., Moldovan, D. I. et al. (2002). Text mining for causal relations. In FLAIRS conference. 360–364.

[19] Miller, G. A. (1995). WordNet: a lexical database for English. Communications of the ACM. 38(11), 39–41.

[20] Jiang, L., Zhang, H. and Cai, Z. (2009). A Novel Bayes Model: Hidden Naive Bayes. IEEE Transactions on Knowledge and Data Engineering. 21(10), 1361–1371. ISSN 1041-4347. doi:10.1109/TKDE.2008.234.

[21] Zeng, D., Liu, K., Lai, S., Zhou, G., Zhao, J. et al. (2014). Relation classification via convolutional deep neural network.

[22] Komninos, A. and Manandhar, S. (2016). Dependency based embeddings for sentence classification tasks. In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. 1490–1500.