One Belt, One Road, One Sentiment? A Hybrid Approach to Gauging Public Opinions on the New Silk Road Initiative

Jonathan Kevin Chandra, Erik Cambria
School of Computer Science and Engineering
Nanyang Technological University
Singapore
jchandra002@e.ntu.edu.sg, cambria@ntu.edu.sg

Andrea Nanetti
School of Art, Design and Media
Nanyang Technological University
Singapore
andrea.nanetti@ntu.edu.sg

***Abstract*—With the rapid adoption of the Internet, fast-moving social media platforms have been able to extract and encapsulate real-time public sentiments on different entities. Real-time sentiment analysis on current dynamic events such as elections, global affairs and sports are essential in the understanding the public's reaction to the states and trajectories of these events. In this paper, we aim to extract the sentiments of the Belt and Road Initiative from Twitter. Using aspect-based sentiment analysis, we were able to obtain the tweet's sentiment polarity on the related aspect category to better understand the topics that were discussed. We have developed an end-to-end sentiment analysis system that collects relevant data from Twitter, processes it and visualizes it on an intuitive display. We employed a hybrid approach of symbolic and sub-symbolic techniques using gated convolutional networks, aspect embeddings and the SenticNet framework to solve the subtasks of aspect category detection and aspect category polarity. A confidence score threshold was used to decide on the results provided by the models from the differing approaches.**

***Index Terms*—Aspect-based sentiment analysis, Twitter sentiment analysis, Deep learning**

I. Introduction

Sentiment analysis is a field of natural language processing (NLP) that aims at extracting subjective information from data [1]. In earlier works, extracting opinion polarity using online reviews of products and services was one of the most common applications of sentiment analysis [2]. Since then, increased computing power and availability of data sources have extended those applications to many domains, including financial forecasting [3], marketing [4], healthcare [5], tourism [6], recommendation systems [7], dialogue systems [8], and more.

It is important for sentiment analysis to be tackled in a holistic manner rather than as separate individual problems. A common term to define the sentiment analysis task is to look at it as a suitcase research problem that involves solving multiple subtasks before opinion extraction is performed [9]. As society moves towards fast-moving and dynamic data sources such as blogs, microblogs (e.g., Twitter), and other social networks, there has been a great interest in how we can use these sources for real-time sentiment analysis applications.

The analysis of dynamic events, such as elections and ongoing sports events, allow for the development of applications that monitor and aggregate real-time public opinions of said events. Extracting the opinions and sentiments towards an entity as a whole is no longer enough. It is also important to look at the opinions made towards different aspects of the entity. Aspect-based sentiment analysis (ABSA) allows us to gain these insights by extracting the aspect category and its related sentiment value of the text.

In this paper, we have designed an integrated end-to-end system that provides us with daily updates of the ABSA of tweets related to the Belt and Road Initiative (BRI). We have developed the data collection, processing and visualization modules required by the system. A hybrid approach of symbolic and sub-symbolic techniques using modified Gated Convolutional Neural Networks (GCNN), modified Gated Convolutional Neural Networks with Aspect Embedding (GCAE) and the SenticNet framework, was used to tackle the two subtasks Aspect Category Detection (ACD) and Aspect Category Polarity (ACP). We have also conducted a case study on the application of this system on new tweets collected at a separate time period. Analysis results from this case study can be used to show the scope of the system and its possible applications for future studies.

A. Research Context: Belt and Road Initiative

Proposed in 2013, China's BRI or One Belt, One Road (OBOR) is one of China's most significant undertaking under President Xi. It focuses on infrastructure building throughout the less developed countries with support from China. Key themes regarding the initiative include China's motivation in terms of economic, energy and security cooperation with member countries [10]. However, concerns regarding safety and environmental hazards of the projects have been highlighted in both traditional and social media. Economic drawbacks affecting member countries, such as debt traps, have also been rising concern. Member countries were unable to pay off their debt and forced to provide other types of concessions and control to China.

A literature review on this issue has highlighted Environment, Energy, Project Operations, Economics and Trade as key themes in academic literature [11]. To gain useful insights on the opinions of the public on the OBOR initiative, it is imperative that the sources of these opinions are subjective and unbiased. Social media may provide us with the opinions of the public. Insights into the sentiment values of the different themes will also give us an overview of the topics that the public is currently concerned about.

II. Related Work

The 'classical' approach to sentiment analysis was to create a model that classifies a positive or negative polarity to input texts [12]. However, as problems get more complex and information becomes more abundant, approaches to these problems have to evolve as well. Sentiment analysis research has advanced through three main paths, namely (1) Granularity: focusing on the size of the target inputs; (2) Extraction Approach: the methods in which sentiments are obtained; and (3) Discourse: using information from text organization [13]. There are generally two main approaches to NLP, namely symbolic and sub-symbolic. The symbolic approach is mainly based on a theory-driven top-down approach using human-made rules and lexicons while the sub-symbolic approach is more data-driven and requires the model to learn from the input text data.

A. Symbolic Approaches to Natural Language Processing

Natural language is inherently a symbolic representation of the human mind and knowledge. Early solutions to tasks involving natural language were symbolic in nature and took advantage of existing linguistic rules to generate expert systems. In such approaches, a text is often considered a collection of words with no relations between individual words (bag-of-words approach). Most of these approaches made use of semantic similarity between words to accomplish their tasks at hand. 'Classical' works include researches where adjective indicators were used to classify subjectivity or sentiment polarity of input texts [14]. Turney [15] was able to predict the sentiment polarity of adjectives and nouns by comparing their similarities to 'excellent' and 'poor'. Lexical databases such as WordNet were a key tool to such approaches, whether to determine the semantic orientation of the word based on similarity relationships between each other [16] or to have used semantic similarity to compare the words to 'good' and 'bad' to determine their sentiments [17].

Knowledge bases have been gaining traction again recently due to the availability of structural information and importance of using commonsense knowledge to create natural language outputs [18]. These databases represent structures and store information on the semantic relationship and structural links between entities, relationships and semantics [19]. Recent applications of NLP using such information include using it for dialogue systems for domain knowledge and commonsense knowledge [20]–[22], fluent natural language generation [23], [24], and complex question answering [25], [26].

B. Sub-symbolic Approaches to Natural Language Processing

Sub-symbolic approaches excel in their ability to process multiple features, accept changes in inputs and have a confidence score which shows the uncertainty of the result despite their lack of interpretability [27]. Deep learning has made impressive advancements in the field of computing and NLP researchers have capitalized on the trend to focus on using such techniques for its applications. Deep learning uses automatic feature extraction which allows for a more effective model training process as compared to previous models which requires hand-crafted features [28]. The concurrent improvements of word embedding techniques [29] and deep learning methods [30] had also led to the skyrocketing of the amount of research in the field. Deep learning models have allowed sub-symbolic approaches to achieve state-of-the-art results for NLP problems with the ability to compute large amounts of data with minimal manual inputs [31]. Ultimately, these approaches still have the issue of interpretability and are based on past observations and, hence, will still have difficulties to produce true natural language solutions. In recent years, many researchers have embedded symbolic lexical databases to sub-symbolic approaches to provide language structure and interpretability to the results. An example includes using role-unbinding vectors to generate sentences and extract partial grammatical structures [32] while another looks into how symbolic representations can be interpreted even within distributed representations such as vectors and tensors [33].

C. Sentic Computing

Sentic computing is an approach that uses concept-level NLP and has an adaptive and content-aware focus on analysis. Unlike traditional NLP techniques, sentic computing does not simply focus on statistical methods and polarity detection. The approach looks at concepts instead of words and uses specific domain knowledge to derive the context of ambiguous texts (e.g., cheap materials having a positive connotation in the context of businesses but having a negative connotation in the context of luxury fashion). Sentic computing uses a three-layer structure, namely syntactic, semantic and pragmatic, and differentiates itself from traditional statistical NLP through three main factors namely: Mono- to Multi-Disciplinarity, Syntax to Semantics, and Statistics to Linguistics [13]. This paradigm shift allows sentic computing to overcome the problem of concept ambiguity and other problems usually associated with statistical approaches. SenticNet is a framework and knowledge base for concept-level sentiment analysis. The framework was initially started by associating a polarity value to about 5,700 concepts mostly obtained from the OpenMind Corpus. However, since then, more updates have been published with increasing improvements in the processing logic. The latest update, SenticNet 6, tackles the issue of new concepts that has does not exist in the knowledge base by using a similarity algorithm and clustering related words together [34].

D. Aspect-Based Sentiment Analysis

Hu and Liu [35] explained that sentiment analysis can be conducted at three levels of granularity: document level, sentence level and aspect level. An analysis of aspect-level sentiment analysis allows the discovery of sentiment values of different terms and aspects within a sentence. An example involves the following sentence: ‘The house has very nice furniture but the cost is steep and it is located in an undesirable location.’ A sentence-level sentiment analysis might not result in an accurate polarity value due to the occurrence of multiple positive and negative sentiments. A more granular aspect-level sentiment analysis shows a positive sentiment with regards to its furniture but a negative sentiment with regards to the house’s cost and location. Most ABSA problems have generally depended on sub-symbolic techniques [36], that is supervised [37], semi-supervised [38] and unsupervised [39] machine learning techniques. There are also multiple subtasks (as shown in Fig. 1) under the umbrella of ABSA as pointed out by the annual SemEval evaluation workshops [40]–[42].

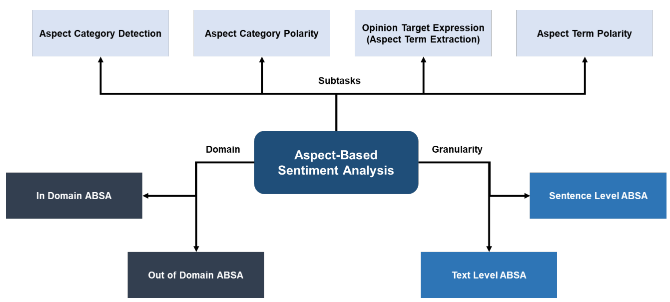


Fig. 1. A brief overview of ABSA subtasks

III. Methodology

Using themes obtained from literature reviews and processing of our collected data, we have chosen four aspect categories where each tweet can be classified into: General, Economic, Environment and Security. We approach this sentiment analysis problem using both symbolic (SenticNet) and sub-symbolic (deep learning) techniques.

A. Data Collection and Preparation

a) Twitter Data: Our Twitter dataset consists of 21,939 tweets posted between 4 September 2019 and 31 January 2020. We scraped and collected these tweets daily using Tweepy [1] and the Twitter API with the following keywords: *Belt and Road Initiative*, *Belt and Road*, *One Belt One Road*, *New Silk Road*, *#OneBeltOneRoad*, *#BeltAndRoadInitiative*, *#BRI*, *#OBOR*, *#NewSilkRoad*. Retweets and duplicated tweets were removed from the dataset. URLs were removed from the tweets. Tweet metadata (Date Posted, Username and Location) were also collected for the dataset.

[1]http://www.tweepy.org/

b) Global Database of Events, Language and Tone (GDELT) Data: GDELT is a real-time open database which collects data on global news media of every country in multiple languages and channels [43]. We obtain daily news event titles relating to the BRI with the purpose of looking at the relationship between Twitter sentiments and global events. Data collected includes Article Title, Date Posted, Language of Article, Country of Origin and URL of the article.

c) Data Cleaning: Data cleaning and preprocessing were performed on the collected Twitter data. We tokenized the raw text into tokens by taking into account phrases that are more commonly used in social media including URLs, speech punctuation, emails, abbreviations, emoticons, decorative characters and embedded apostrophes. Tweets below 10 tokens were removed to ensure the tweets are long enough to convey enough meaning. Mentions and hyperlinks were removed and the hex character (#) was removed from the hashtags. Multiple consecutive duplicate tokens were also removed. Duplicate tweets were then removed from the resulting dataset. After preprocessing, 14,397 unique tweets remain.

d) Preparation of Gold Standard Tweets: From the preprocessed tweets, we segregated the tweets made from 4 September 2019 to 13 December 2019 for both our training and testing dataset. From the 9,004 tweets segregated for the training and testing dataset, 950 unambiguous gold standard tweets were selected and annotated into one of the four aspect categories. Tweets were classified to their respective categories based on the following criteria.

- **Economic:** Discussions on the economic implications towards society, member countries, individuals or the global economy with references to debt, investments, poverty and money.
- **Environment:** Discussions on the environment including renewable energy, pollution, environmental impacts and efforts towards green energy.
- **Security:** Discussions on the explicit or implicit mentions of war, slavery, torture, security, privacy, sovereignty and other security implications.
- **General:** Discussions related to the BRI that have not been classified into the other aspect categories.

The gold standard tweets should be unambiguous and were selected based on the confidence level of the annotators to classify them into one of the four aspect categories. For these tweets, the inter-annotator score was found to be 81.3% for aspect category and 87.6% for sentiment polarity.

e) Word Embeddings: Following the successes of SemEval ABSA submissions, we propose to use Global Vectors for Word Representation (GloVe) as our word embeddings for the task of ABSA. We used the Common Crawl GloVe version [44], a pre-trained 300-dimension vector representation database of 840 billion tokens and 2.2 million vocabulary, to convert our preprocessed tweets into word embeddings.

B. Data Processing: Deep Learning

a) ACD using Modified GCNN: For the subtask of ACD, we present a modified GCNN originally presented as a model

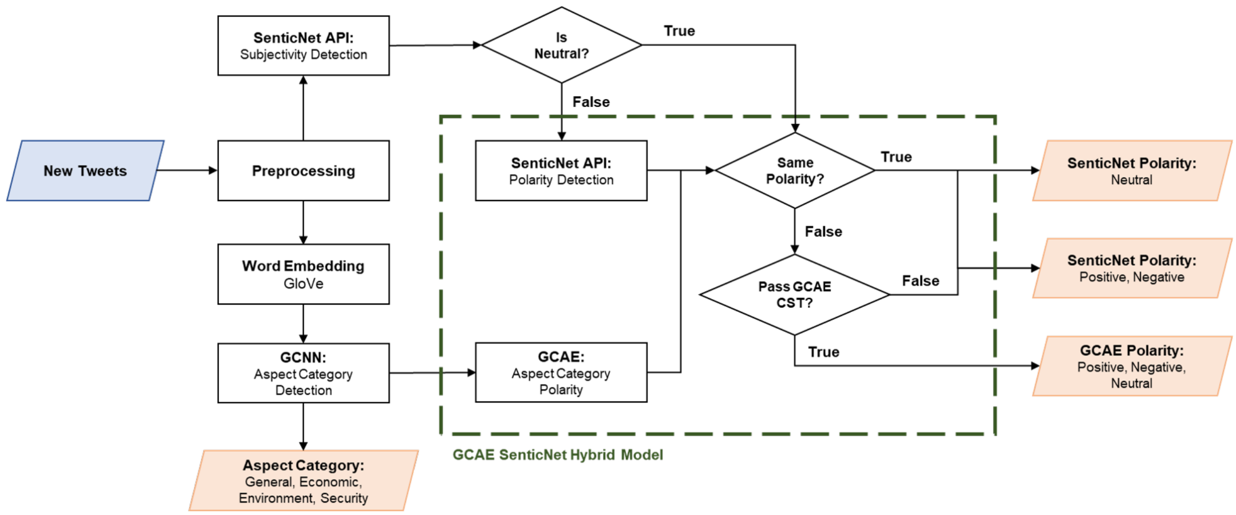


Fig. 2. One Belt One Road One Sentiment Processing Flow Chart for New Tweets

used for aspect category sentiment analysis tasks [45]. The convolutional layer and max pooling layer use a similar architecture as in a vanilla Convolutional Neural Network (CNN) for text classification [46]. A modified GCNN model architecture was used to classify our preprocessed tweets into one of the four aspect categories. This model includes a one-dimensional convolutional layer, a pair of gating mechanisms, a max pooling layer and a softmax layer. Gated Tanh and Linear units have been shown as effective gating mechanisms in language modeling tasks. The gated convolutional filters have the following equations:

$$a_i = relu(X_{i:i+k} * W_a + b_a) \tag{1}$$

$$s_i = tanh(X_{i:i+k} * W_s + b_s) \tag{2}$$

$$c_i = s_i \times a_i \tag{3}$$

Where W_i represents the convolutional filter, * represents the convolution operation, b_i represents the bias, and $\times$ represents the element-wise multiplication operation. The max pooling layer obtains the maximal value among the convolutional features $c = [c_1, c_2, , c_L]$ and generate a fixed size filter $e \in R^{d_k}$. The softmax layer uses the vector e to predict its aspect category $\hat{y}$.

As compared to other deep learning algorithms, CNNs aim to extract the most essential n-grams from the text to perform classifications while generally ignoring long-term dependencies. This model performs best by taking advantage of coarse-grained local and deep features of the text [47]. We believe this model would be suitable to process short texts for this paper.

GCNN is also less complex and more efficient than recurrent network-based models [48] [49]. GCNN uses convolutional layers and gating units which are independent from each other. This allows the model to be computed in parallel and takes advantage of modern computing and hardware to achieve a more efficient result.

b) ACP using Modified GCAE: For the subtask of ACP, we present a modified GCAE originally presented as a model used for aspect category sentiment analysis tasks [45]. Aspect category obtained from GCNN is converted to aspect embeddings and be inputted into our modified GCAE to obtain a polarity classification. The inclusion of the aspect embedding allows the aspect category of the text to be considered when predicting for its polarity value. N-grams of text from different aspect categories may have different sentiment values. For example, the phrase 'building more coal power plants' may have a positive sentiment value for the 'Economic' aspect category while having a negative sentiment value for the 'Environment' aspect category. Additionally, the phrase 'economic advancements' may have a greater positive sentiment value in the 'Economic' aspect category than the equivalent clause in a 'General' aspect category. The modified GCAE model is similar to the modified GCNN model with the exception of adding an extra aspect category embeddings component into the ReLU gate and outputting a sentiment polarity value instead of an aspect category value. The modified gated convolutional filter a_i have the following equation:

$$a_i = relu(X_{i:i+k} * W_a + V_a v_a + b_a) \tag{4}$$

Where $V_a v_a$ represents the additional aspect embedding information. The Tanh gate is responsible for computing the sentiment features while the ReLU gate is responsible for the aspect features. The results of these two gates are multiplied element-wise and inputted into the max pooling layer. A fixed-size vector and logit are generated in the max pooling layer that extracts the most important features of the text. Softmax layer then produces the output sentiment polarity of the text.

c) Experiment Settings: The model is trained by minimizing the cross-entropy loss between the actual value y and predicted value $haty$ for all training samples i and aspect category classes j with the following loss equation:

$$\mathcal{L} = -\sum_i \sum_j y_i^j \lg \hat{y}_i^j \quad (5)$$

Most of the hyperparameters were fixed for this experiment. Learning rate was set at 0.01 and SGD momentum was fixed at 0.99. For the GCNN and GCAE, the probability of a dropout was set at 0.5 and the number of kernels as 100 for each size 3, 4 and 5. Our training and evaluation sequence did not use CUDA for this experiment. We use a 80/20 train/evaluation ratio to split the annotated Gold Standard Tweets using stratified sampling. The distribution of the dataset are shown in Table I and Table II.

TABLE I
DATASET DISTRIBUTION BY ASPECT CATEGORY

Aspect Category	Total	Train	%	Eval	%
General	261	209	80	52	20
Economic	273	217	79	56	21
Environment	209	171	82	38	18
Security	207	164	79	43	21

TABLE II
DATASET DISTRIBUTION BY SENTIMENT POLARITY

Sentiment Polarity	Total	Train	%	Eval	%
Positive	218	179	82	39	18
Negative	460	371	81	89	19
Neutral	272	211	78	61	22

C. Data Processing: SenticNet

a) Polarity Detection: We use the SenticNet API to perform polarity detection on our tweets. SenticNet allows for the classification of the overall sentiment polarity value of the input sentence using a concept-based sentiment analysis technique. The framework uses a hybrid approach which includes the use of semantic networks, conceptual dependency representations, deep neural networks and multiple kernel learning to tackle the task of concept-based sentiment analysis. The framework processes the input sentence through multiple segments before a polarity is extracted. The sentence goes through a Sentic Parser to pre-process and tokenize the sentence into a bag-of-concepts before being checked against the SenticNet concept database [34]. If the concepts or sentic patterns are not in the concept database, the sentence will be treated as a bag-of-words and will be processed by NLP-Capsule, a state-of-the-art capsule network used to model hierarchical relationships, to output polarity result [50].

Sentic patterns are linguistic patterns that considers the syntactic dependency relations between the concepts in the input sentence [51]. If these patterns are found in the input sentence, the sentiment of the input sentence will be processed concept by concept using dependency-based rules and will result in an overall polarity value. If sentic patterns are not found, the input sentence will be passed through Sentic LSTM [52], a sentiment-enriched LSTM, to obtain a sentiment polarity value.

b) Subjectivity Detection: SenticNet API is also used for subjectivity detection on the tweets. It uses a lexical affinity technique to assign probabilistic affinity for different categories of word. Similar to the SenticNet Polarity Detection module, the Sentic Parser is used to break down the text into clauses and concepts. To find a polarity value, the deconstructed concepts are classified into a specific polarity value using AffectiveSpace, a vector space representation of affective commonsense knowledge. However, AffectiveSpace is unable to extract objective opinions from subjective opinions. To solve this, a deep learning model is trained and used to identify objective texts in the dataset by taking into account the position of the concept in AffectiveSpace.

D. Data Processing: Hybrid Model for Polarity Detection using Modified GCAE and SenticNet

For this paper, we propose an additional hybrid approach for polarity detection by using both the modified GCAE model and SenticNet. Both models are first used to classify the sentiment polarity. If both models produce the same classification, there is no issue. If there is a conflict, check if the confidence score of the GCAE is above the Confidence Score Threshold (CST). If it is above the threshold, the value given by GCAE is used, else, the value given by SenticNet is used. The modified GCAE confidence score is calculated based on the logit of the prediction. The logit is the probability score mapped to $[-\infty, \infty]$ that is given to the softmax layer where the class with the highest probability is outputted. We use this logit score to determine the confidence of the model in its prediction.

E. Visualization

The results from the modules above have to be visualized to give us a more intuitive view on the insights regarding their distribution, word frequency, and change over time. For the visualization module, we designed three types of charts with their respective purposes.

- **Distribution over time:** used to visualize in the distribution of sentiments, aspects or tweet activity over a given time period. Visualized using a line chart
- **Overall distribution:** used to visualize the overall distribution (sentiments or aspects) of the collected tweets as a whole or by aspect.
- **Word cloud:** used to visualize the word frequency of the collected tweets as a whole or by aspect category to give us the distribution of words for each topic.

Future work may incorporate more types of visualization and also introduce interactivity to the system.

IV. RESULTS

A. Metrics

The results of the models were evaluated based on the Macro-F1 and Test Accuracy score. The metrics were computed as follows where N is the number of classes:

$$F1 = 2 * \frac{precision * recall}{precision + recall} \quad (6)$$

$$Macro - F1 = \frac{\sum_n^N F1(n)}{N} \quad (7)$$

$$Test\ Accuracy = \frac{number\ of\ true\ positives}{total\ number\ of\ instances} \quad (8)$$

B. Aspect Category Detection using Modified GCNN

The performance of our modified GCNN model is shown in Table III. Our model produced satisfactory F1-scores for ACD with an F1-score of above 70% for all aspect categories. We see a possible correlation between the breadth of topics being discussed for each aspect category and accuracy of the predicted class. For example, for aspect category 'Environment', a high F1-score can be observed as a majority of the discussed topics revolve around the same few topics which includes discussion regarding the availability or lack of energy sources (e.g., coal, wind, solar) and environmental impacts of the initiative on the marine ecosystem. However, for the aspect category 'Security', discussed topics range widely from human trafficking and slave brides to drone spying and military movements. There may be some difficulty in classifying unseen security-related topics as either "Security" or "General".

TABLE III
PERFORMANCE OF OUR MODIFIED GCNN MODEL ON THE ACD SUBTASK

Aspect Category	Recall	Precision	F1-Score
General	78.9%	70.7%	74.6%
Economic	82.1%	79.3%	80.7%
Environment	97.4%	94.9%	96.1%
Security	65.1%	82.4%	72.7%
Macro-F1	**81.0%**		
Test Accuracy	**80.0%**		

C. Aspect Category Polarity using Hybrid Model

To determine the CST, we analyzed the results found in the evaluation procedure. There is a need to balance both a high enough test accuracy while at the same time have a low enough threshold that the results from the deep learning model is still able to be used for most predictions. If the threshold is set too high, most predictions made by the model will not be used as the confidence score will be too low to be included. Arbitrary thresholds were set from 0 to 9 and the accuracy of the model and the percentage of predictions included were collected and shown in Table IV.

TABLE IV
TEST ACCURACY OF OUR MODIFIED GCAE BY CST

Confidence Score Threshold	Test Accuracy	Predictions Included
0	65.8%	98.9%
1	65.7%	91.0%
2	69.2%	75.7%
3	**71.4%**	**59.3%**
4	79.1%	35.4%
5	86.0%	22.8%
6	85.2%	14.3%

We chose the CST of the modified GCAE model to be 3. From the table shown, this threshold can provide us with at least 70% test accuracy while more than half of its predictions were still included in the analysis. By using a combination of both approaches, we were able to capture both in-domain and out-of-domain information in our model. Table V and Table VI show the performance of our Hybrid Model on the ACP subtask for each class with and without the CST respectively. Our model performed with an improved test accuracy and F1-score with the addition of the CST.

TABLE V
PERFORMANCE OF OUR MODEL ON ACP WITH NO RESTRICTIONS

Sentiment Polarity	Recall	Precision	F1-Score
Positive	43.6%	60.7%	50.7%
Negative	83.1%	74.7%	78.7%
Neutral	55.7%	54.8%	55.3%
Macro-F1	**61.6%**		
Test Accuracy	**66.1%**		

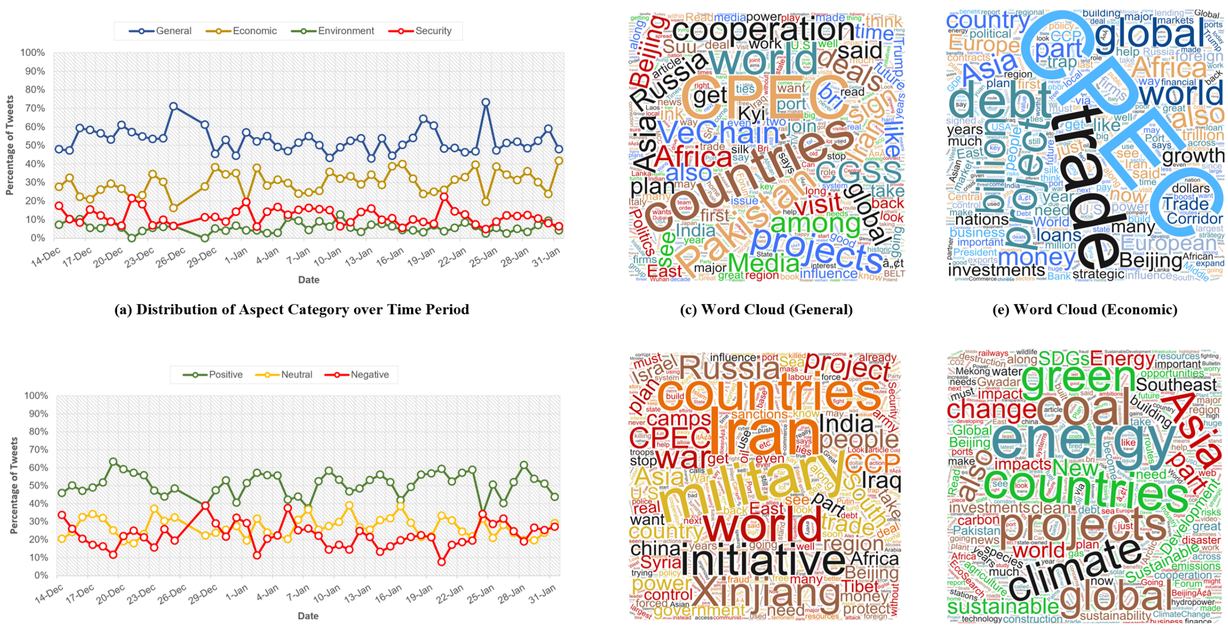
TABLE VI
PERFORMANCE OF OUR MODEL ON THE ACP SUBTASK WITH CST

Sentiment Polarity	Recall	Precision	F1-Score
Positive	52.6%	58.8%	55.6%
Negative	88.3%	77.9%	82.8%
Neutral	51.5%	63.0%	56.7%
Macro-F1	**65.0%**		
Test Accuracy	**71.4%**		

A possible explanation for this difference includes the difficulty of understanding context in short texts without extra processing. For example, "Military drone continuously observing the Belt and Road projects in member countries" is ambiguous without its background context. The tweet can either mean an improvement in security denoting a positive sentiment or an invasion of privacy denoting a negative sentiment. Another possible explanation is the frequency of sarcasm and rhetorical questions which is difficult to classify into either polar or neutral categories. For example, "Belt and Road - an environmentally friendly initiative?" denotes a neutral polarity due to its intention to start a discussion but would be wrongly classified as positive. Another example, "Amazing, my privacy is well protected with all these drones and cameras spying on me #BRI" may denote a negative polarity as the author sarcastically announced his discomfort of being spied upon but would be classified as positive polarity without an extra layer of sarcasm detection.

V. CASE STUDY

To illustrate the application of this paper, we have extracted and processed new tweets for the time period of 14 December 2019 to 31 January 2020 for analysis. Results and analysis regarding the tweets during this time period will be discussed in this segment. A total of 5,390 related tweets were extracted and preprocessed. The following charts shows the analysis results and distribution of the processed tweets. For word frequencies, searched keywords such as *OBOR*, *One Belt One Road* and *New Silk Road* were removed.



(a) Distribution of Aspect Category over Time Period

(b) Distribution of Sentiment Polarity over Time Period

(c) Word Cloud (General)

(d) Word Cloud (Security)

(e) Word Cloud (Economic)

(f) Word Cloud (Environment)

Fig. 3. Visualization for Distribution of Tweets and Word Frequency through Case Study Time Period

A. Overall Distribution of Tweets over Time

From Fig. 3, we can observe that the majority of the tweets were classified in the ‘General’ category followed by ‘Economic’, ‘Security’ and ‘Environment’. ‘General’ category is expected to be the largest as many other different topics or categories that we have not defined were classified in this category. The proportions of the other categories were also in line with expectations based on existing literature review.

Regarding sentiment polarity, it was a surprise that positive sentiments was dominant among the extracted tweets. Upon closer inspection of the tweets extracted, we observed that some tweets that were classified as positive were sarcastic or rhetorical questions. Examples of such tweets include “Wow China is so perfect that can’t even produce their own healthy formula milk for their children nice good one belt one road for developing country” and “Who says CCP wont try to conquer the world?” Such tweets should be classified as either neutral or negative but would ultimately depend on the context of the user, which is difficult to capture in short texts such as tweets.

B. Word Cloud

The charts in Fig. 3 also show the word frequencies of the tweets for each aspect category. The following are some insights for each aspect category:

- **General:** Potential new trending aspect categories could be observed in this category. An example is ‘VeChain’, which is a BRI-related blockchain-enabled platform.
- **Economic:** Expected economic-related words were found here such as ‘trade’, ‘investment’, ‘infrastructure’, ‘debt’, and ‘debt trap’.
- **Environment:** BRI tweets related to this category revolve around two themes namely ‘energy’ and ‘sustainable development’ with positive.
- **Security:**This category is relatively broad and includes themes such as ‘war’, ‘oppression’, ‘power’, ‘control’, and ‘safety’. Tweets also mostly involve countries such as ‘Iran’, ‘Russia’, and ‘Iraq’.

VI. Conclusion and Future Work

In this paper, we have developed a sentiment analysis system that performs ABSA on tweets related to the BRI. Future work may include (1) improving data collection and preparation through other sentiment analysis processes such as microtext analysis and sarcasm detection; (2) implementing a dynamic aspect category system that adds and removes new aspect categories based on current trends; (3) incorporating pre-trained language models (e.g., BERT) into the system; (4) implementing an interactive dashboard with real-time updates on the display to encourage widespread use; and (5) benchmarking with other state-of-the-art models for better implementation.

References

[1] E. Cambria, H. Wang, and B. White, “Guest Editorial: Big Social Data Analysis,” Knowledge-Based Systems, pp. 1-2, 2014.

[2] E. Cambria, T. Mazzocco, A. Hussain, and C. Eckl, “Sentic Medoids: Organizing Affective Common Sense Knowledge in a Multi-Dimensional Vector Space,” Advances in Neural Networks, pp. 601-610, 2011.

[3] A. Picasso, S. Merello, Y. Ma, L. Oneto, and E. Cambria, “Technical analysis and sentiment embeddings for market trend prediction,” Expert Systems with Applications, vol. 135, pp. 60-70, 2019.

[4] E. Cambria, M. Grassi, A. Hussain, and C. Havasi, “Sentic computing for social media marketing,” Multimedia Tools and Applications, vol. 59, no. 2, pp. 557-577, 2012.

[5] E. Cambria, A. Hussain, T. Durrani, C. Havasi, C. Eckl, and J. Munro, "Sentic computing for patient centered applications," in IEEE International Conference on Signal Processing, 2010, pp. 1279-1282.

[6] Y Li, S Wang, Y Ma, Q Pan, E Cambria, "Popularity Prediction on Vacation Rental Websites," Neurocomputing 412, pp. 372-380, 2020.

[7] Y Li, S Wang, Q Pan, H Peng, T Yang, E Cambria, "Learning Binary Codes with Neural Collaborative Filtering for Efficient Recommendation Systems," Knowledge-Based Systems 172, pp. 64-75, 2020.

[8] Y. Ma, K.L. Nguyen, F. Xing, and E. Cambria, "A survey on empathetic dialogue systems," Information Fusion, vol. 64, pp. 50-70, 2020.

[9] E. Cambria, S. Poria, A. Gelbukh and M. Thelwall, "Sentiment Analysis Is a Big Suitcase," Affective Computing and Sentiment Analysis, pp. 74-80, 2017.

[10] K.-J. Yii, K.-Y. Bee, W.-Y. Cheam, Y.-L. Chong and C.-M. Lee, "Is Transportation Infrastructure Important to the One Belt One Road (OBOR) Initiative? Empirical Evidence from the Selected Asian Countries," MDPI Sustainability Vol. 10, 2018.

[11] W. Li and O.-P. Hilmola, "One Belt One Road: Literature Analysis," Transport and Telecommunication Vol. 20 No. 3, pp. 260-268, 2019.

[12] B. Pang and L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales," Proceedings of the ACL, 2005.

[13] E. Cambria, A. Hussain, C. Havasi, and C. Eckl, "Sentic computing: Exploitation of common sense for the development of emotion-sensitive systems," ser. LNCS. Springer, 2010, vol. 5967, pp. 148-156.

[14] V. Hatzivassiloglou and J. M. Wiebe, "Effects of Adjective Orientation and Gradability on Sentence Subjectivity," in COLING 2000 Volume 1: The 18th International Conference on Computational Linguistics, 2000.

[15] P. D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," in 40th Annual Meeting of the ACL, Philadelphia, 2002.

[16] J. Kamps and M. Marx, "Words with attitude," in Proceedings of the 1st International Conference on Global WordNet, Mysore, India, 2002.

[17] A. Budanitsky and G. Hirst, "Evaluating WordNet-based Measures of Lexical Semantic Relatedness," in Computational Linguistics, Volume 32, Number 1, 2006.

[18] E. Cambria, A. Hussain, C. Havasi, and C. Eckl, "Common Sense Computing: From the Society of Mind to Digital Intuition and Beyond," Biometric ID Management and Multimodal Communication, pp. 252-259, 2009.

[19] S. Ji, S. Pan, E. Cambria, P. Marttinen and P. S. Yu, "A Survey on Knowledge Graphs: Representation, Acquisition and Applications," arXiv preprint arXiv:2002.00388, 2020.

[20] Y. Ma, P. A. Crook, R. Sarikaya and E. Fosler-Lussier, "Knowledge Graph Inference for spoken dialog systems," in IEEE International Conference on Acoustics, Speech and Signal Processing, Brisbane, QLD, Australia, 2015.

[21] B. Hixon, P. Clark and H. Hajishirzi, "Learning Knowledge Graphs for Question Answering through Conversational Dialog," in Conference of the North American Chapter of the ACL: Human Language Technologies, 2015.

[22] A. Saha, V. Pahuja, M. M. Khapra, K. Sankaranarayanan and S. Chandar, "Complex Sequential Question Answering: Towards Learning to Converse Over Linked Question Answer Pairs with a Knowledge Graph," in Thirty-Second AAAI Conference on AI, 2018.

[23] R. Logan, N. F. Liu, M. E. Peters, M. Gardner and S. Singh, "Barack's Wife Hillary: Using Knowledge Graphs for Fact-Aware Language Modeling," in 57th Annual Meeting of the ACL, Florence, Italy, 2019.

[24] A. Moryossef, Y. Goldberg and I. Dagan, "Step-by-Step: Separating Planning from Realization in Neural Data-to-Text Generation," CoRR, 2019.

[25] A. Saha, G. A. Ansari, A. Laddha, K. Sankaranarayanan and S. Chakrabarti, "Complex Program Induction for Querying Knowledge Bases in the Absence of Gold Programs," Transactions of the Association for Computational Linguistics Vol 7, p. 185200, 2019.

[26] L. Weber, P. Minervini, J. Mnchmeyer, U. Leser and T. Rocktschel, "NLProlog: Reasoning with Weak Unification for Question Answering in Natural Language," 2019.

[27] E. Boiy and M.-F. Moens, "A Machine Learning Approach to Sentiment Analysis in Multilingual Web Texts," Journal of Information Retrieval, pp. 526-558, 2009.

[28] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, "Deep Learning based Text Classification: A Comprehensive Review," ACM Computing Surveys 54, 2021.

[29] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," in 26th International Conference on Neural Information Processing Systems Volume 2, 2013.

[30] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in Proceedings of the conference on empirical methods in natural language processing (EMNLP), Seattle, 2013.

[31] LeCun, Bengio and H. G, "Deep learning.," Nature, pp. 436-444, 2015.

[32] Q. Huang, P. Smolensky, X. He, L. Deng and D. Wu, "A Neural-Symbolic Approach to Natural Language Tasks," 2018.

[33] L. Ferrone and F. M. Zanzotto, "Symbolic, Distributed and Distributional Representations for Natural Language Processing in the Era of Deep Learning" 2019.

[34] E. Cambria, Y. Li, F. Xing, S. Poria, and K. Kwok, "SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis," in CIKM, pp. 105-114, 2020.

[35] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, 2004.

[36] H. H. Do, P. Prasad, A. Maag and A. Alsadoon, "Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review," Expert Systems with Applications Vol 118, pp. 272-299, 2019.

[37] I. Chaturvedi, E. Ragusa, P. Gastaldo, R. Zunino, and E. Cambria, "Bayesian network based extreme learning machine for subjectivity detection," Journal of The Franklin Institute, vol. 355, no. 4, pp. 1780-1797, 2018.

[38] A. Hussain and E. Cambria, "Semi-supervised learning for big social data analysis," Neurocomputing, vol. 275, pp. 1662-1673, 2018.

[39] N. Ofek, S. Poria, L. Rokach, E. Cambria, A. Hussain, and A. Shabtai, "Unsupervised commonsense knowledge enrichment for domain-specific sentiment analysis," Cognitive Computation, vol. 8, no. 3, pp. 467-477, 2016.

[40] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos and S. Manandhar, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis," in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, 2014.

[41] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar and I. Androutsopoulos, "SemEval-2015 Task 12: Aspect Based Sentiment Analysis," in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, Colorado, 2015.

[42] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos and e. a. Suresh Manandhar, "SemEval-2016 Task 5: Aspect Based Sentiment Analysis," in Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), San Diego, California, 2016.

[43] P. A. S. Kalev Leetaru, "GDELT: Global data on events, location, and tone," in ISA Annual Convention, 2013.

[44] J. Pennington, R. Socher and C. D. Manning, "GloVe: Global Vectors for Word Representation," 2014.

[45] W. Xue and T. Li, "Aspect Based Sentiment Analysis with Gated Convolutional Networks," in Proceedings of the 56th Annual Meeting of the ACL, Melbourne, 2018.

[46] Y. Kim, "Convolutional Neural Networks for Sentence Classification," EMNLP, pp. 1746-1751, 2014.

[47] W. Yin, K. Kann, M. Yu and H. Schtze, "Comparative Study of CNN and RNN for Natural Language Processing," arXiv:1702.01923, 2017.

[48] Y. Wang, M. Huang, X. Zhu and L. Zhao, "Attention-based LSTM for Aspect-level Sentiment Classification," in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, Texas, 2016.

[49] D. Ma, S. Li, X. Zhang and H. Wang, "Interactive Attention Networks for Aspect-Level Sentiment Classification," in Interactive Attention Networks for Aspect-Level Sentiment Classification, 2017.

[50] W. Zhao, H. Peng, S. Eger, E. Cambria and M. Yang, "Towards Scalable and Reliable Capsule Networks for Challenging NLP Applications," in Proceedings of the 57th Annual Meeting of the ACL, pp. 1549-1559, 2019.

[51] S. Poria, E. Cambria, G. Winterstein and G.-B. Huang, "Sentic patterns: Dependency-based rules for concept-level sentiment analysis," Knowledge-Based Systems Vol 69, pp. 45-63, 2014.

[52] Y. Ma, H. Peng, E. Cambria, T. Khan and A. Hussain, "Sentic LSTM: a Hybrid Network for Targeted Aspect-Based Sentiment Analysis," Cognitive Computation Vol 10 Issue 4, pp. 639-650, 2018.