The Narrative Fallacy in News Coverage of China's Belt and Road Initiative

SHANTANU SHARMA

U1622895F

Supervisor: Dr. Anwittaman Datta

SCSE19-0208

Submitted in Partial Fulfillment of the Requirements for the Degree of Bachelor of Computer Science at Nanyang Technological University, Singapore

Abstract

The last two decades connected humanity on a never-seen-before level. Along obvious social and psychological: benefits and detriment arising through rise of connectivity, there was a clandestine and subversive aftermath. Black propaganda is a form of propaganda intended to create the impression that it was created by those it is supposed to discredit. It is typically used to vilify, embarrass, or misrepresent the enemy and mold public opinion. The term fake news, as popularized by Donald Trump, is a subset of such operations. Through analysis of data collected, we would like to identify fidelity of media towards the Belt and Road Initiative (BRI) , a multi-trillion dollar project announced by the Chinese government in 2013, and identify if there is a biased sentiment in western media compared to other media outlets. The project contributes a new approach for pre-processing of news data for sentiment analysis and uses data analysis to prove the bias hypothesized.

Acknowledgement

I would like to take this opportunity to express my deep gratitude to my Final Year Project Supervisor, **Professor Anwitaman Datta,** for his constant guidance, lessons and patience throughout the project. He was extremely supportive and was able to guide me to pursue the topic. He truly helped me understand the value of effective communication and discipline.

I would like to thank Vansh, Chaitanya, Shourya, and Puneet for providing me the support throughout the endeavour. I am also grateful to my brother Ishank and my family for providing motivation through their kind and disparaging remarks.

I also want to acknowledge my companions Vinith, Cybele, Vansh, Aish, Kavya, Ron, Mahir, Saranya, Lisa, Kirti, Delhi clique, and Ifor Cult who made me who I am today.

“You know, the Chinese say that once you've saved a person's life, you're responsible for it forever.”

-John "Scottie" Ferguson, Vertigo

Contents

List of figures

Chapter 1 **Introduction**

1.1 **Motivation**

A common household tale of food product- MSG has pervaded communities throughout the world. Symptoms caused by MSG range from headache, indigestion, severe vomiting to weak bones. “Chinese Restaurant Syndrome” is a quintessential example of misleading propaganda [1,2]. “ *Dezinformatsiya*”, French-sounding word claiming to have a Western origin, was coined by Joseph Stalin and derived from the title of a KGB black propaganda department [3,4]. The English word disinformation is a loan translation. Operation INFEKTION, a Soviet disinformation campaign to influence opinion that the U.S. invented AIDS started from a news article published in an less-known Indian newspaper caused a butterfly effect and ended up being covered by Dan Rather on primetime American television.

Even small communities with little or no financial and technical resources like online forum 4chan have been able to cause rippling effects throughout the globe ranging from: spreading “#BaldforBieber”, an internet hoax claiming Justin Bieber had been diagnosed with cancer resulting in several teen shaving their heads in solidarity; to disinformation with malicious intent and profound consequences like Pizzagate scandal resulting in a citizen firing a rifle inside the Comet Ping Pong restaurant [5].

These above mentioned occurrences provided motivation to analyze development of news stories and whether biases existing in the media can be propagated through false claims. As mentioned above, biased/ falsified/ incomplete new coverage can lead to adverse consequences, in this project we investigate media coverage of Belt and Road Initiative (BRI) by China and structural risks of global trade which can result as a consequence. Although it is arduous to analyze the direct economic and social impact of news stories, this project attempts to identify the changing narrative about Belt and Road Initiative and would try to confront the reality based upon existing research, data gathered, policy consideration, and past databases about this venture.

On his overseas official trip in September and October 2013, President Xi Jinping addressing diplomats and students at university in Kazakhstan announced the project for overland routes for road and rail transportation, called "the Silk Road Economic Belt”. Few days later in Indonesia, the president announced the project called “21st Century Maritime Silk Road” emphasising on reviving centuries old ties with countries and promoting trade of commodities and culture- "a bid to enhance regional connectivity and embrace a brighter future". The combined project was titled “One Belt, One Road” (OBOR) and was also known as New Silk Road Project. The project was later renamed to BRI. The project encompasses infrastructure investment, education, construction materials, railway and highway, automobile, real estate, power grid, and iron and steel. BRI has grown since its announcement to include activities in the Arctic, cyberspace, and even outer space. As of April 2019, 125 countries and 29 international organizations have signed BRI cooperation documents with China [42] which covers 75%+ of the world's population and 52%+ global GDP.

The project since its inception has been a subject of controversy and has been called a ploy for **“debt-trap diplomacy”** where in China loans large sums of money (typically in billions of dollars) to countries with weak economies and risk of default. These countries not adept to pay-back coupon/principal payment eventually have to surrender their assets. Intriguingly these countries have engaged in more deals and borrowed further loans from China which hinted a fissure between reality and the media-sphere.

1.2 Objective

The project aims to investigate “debt-trap” indoctrination of Chinese loans offered under BRI and analyze the legitimacy of the “debt-trap” claim. The following project is carried in phases with corresponding direction-

Phase 1: Perform web-scraping and crawling of news websites to create a database of news articles mentioning China’s BRI or debt-trap diplomacy. Analyze data and mark data for opinion within strict defined rules.
Phase 2: Performing sentiment analysis on data gathered. Improve accuracy of sentiment model for opinion analysis on news articles
Phase 3: Analyze database constructed to support claim for polarity in media coverage.
Phase 4: Using existing literature and policy review from historical data to uncover which camp has the most objective view.

1.3 Contributions

The project was successfully completed, with all its objectives being met. The main contributions of the project can be listed down as follows:

1. Performed sentiment analysis and was able to propose a pre-processing technique to enhance accuracy which is tailored to analysis in news articles.
2. Corroborated the claim for media bias via data analysis and policy literature review

1.4 Report Organisation and Structure

This report is organized as follows:

• Chapter 2 aims at providing necessary background information relevant to the project including web-scraping, Convolution Neural Networks, LSTMs, Word Embedding as preliminaries.
• Chapter 3 comprises survey of current literature.
• Chapter 4 provides relevant information about the datasets used and strategies and principles employed.
• Chapter 5 discusses the experiments performed and the methods used to implement them. The results obtained in the experiments are also analyzed.
• Chapter 6 analyzes the result further and look at case studies
• Chapter 7 includes some concluding remarks and statements and discusses what potential future work could be done in this project.

Chapter 2 **Preliminaries**

In this section we will emphasize on illustrating our choices methods and give a detailed explanation on the general methods that are utilised in later sections.

2.1 **Web-scraping and news extraction**

News articles can be used for research purposes for creating time-series analysis of change in information, to track source of data. There are many freely available dataset and databases online such as RCV1 [6]. News article scraping is done using analysing RSS of a website while crawling is done using recursive hyperlink exploration. Most web-scraping apis in python are built upon *request* library.

2.2 **Artificial Neural Networks**

"Artificial neural networks" (ANN) or "Neural networks" (NN) is a class of pattern recognition systems inspired by the design of the human brain. ANNs have been widely successful in modeling classification and regression networks. The basic building block of an artificial neural network is an artificial neuron. Each artificial neuron in a network acts as a node similar to neurons constituting the nervous system of the human body. Decision of which message to pass to the next neuron is considered via activation function. The most prominent activation function is Rectified Linear Unit Activation Function (ReLU).

2.2.1 ReLU Activation Function

One of the most common activation functions used is Rectified Linear Unit Activation Function (ReLU). ReLU activation function and artificial neuron relation leads to neural networks.

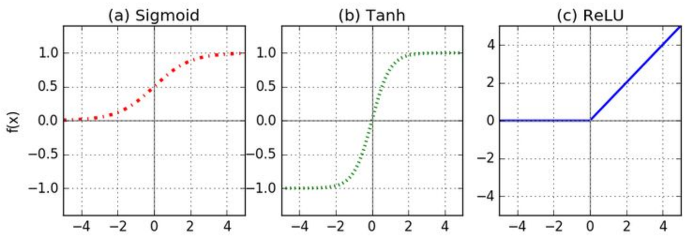


Figure 1: The output behaviour of different activation functions.

2.2.2. Multi-Layer Neural Network and Deep Neural Network

Many deep learning architectures have proven to outperform humans in several structured learning tasks. Such network architectures are referred to as Deep Neural Networks or DNNs. A neural network consists of connected neurons. A node is a computational representation of biological neurons and connections in the human brain. Each node is defined by an activation function and a collection of such nodes is called a hidden layer. In DNN models, there are several such neuron layers stacked together. The performance of a model can be calculated using a loss function. In general, DNNs can be trained with unsupervised and supervised learning techniques. In supervised learning, we use labeled data to train the DNNs and learn the weights that minimize the loss function for classification or regression tasks [43]. Refer Figure 2 for an illustration of a Deep Neural Network.

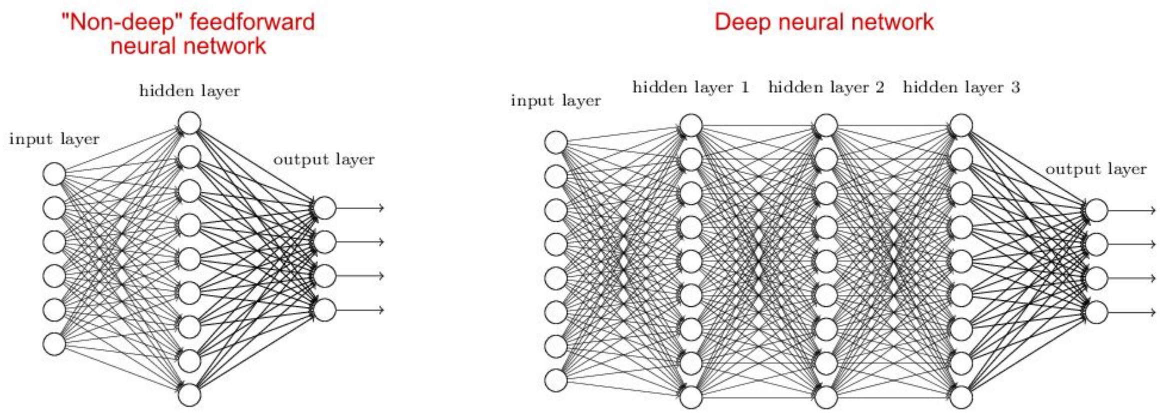


Figure 2: Illustration of a Deep Neural Network. Adapted from [44]

2.2.4 Recurrent Neural Network

A recurrent neural network (RNN) is a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence. This allows it to exhibit temporal dynamic behavior. Derived from feedforward neural networks, RNNs can use their internal state (memory) to process variable length sequences of inputs. It can take one or more input vectors and is useful in cases with sequential arrangement that can be identified. The new predicted values are influenced not just by weights but also by a "hidden" state vector that represents a context based on the type of RNN model being used. Major challenges faced by RNN are vanishing and exploding gradients which is a result of too small or too large weights accordingly leading to elimination or distraction of value learned. Such results occur especially in the case of a large number of states.

2.2.5 Convolution Neural Network

One of the most popular deep learning models is Convolution Neural Network (CNN). CNN can help tackle problems faced in RNN, since there is no timestamp in case of CNN, the processes can be performed in parallel and can be done much faster than the time taken to train an RNN model.

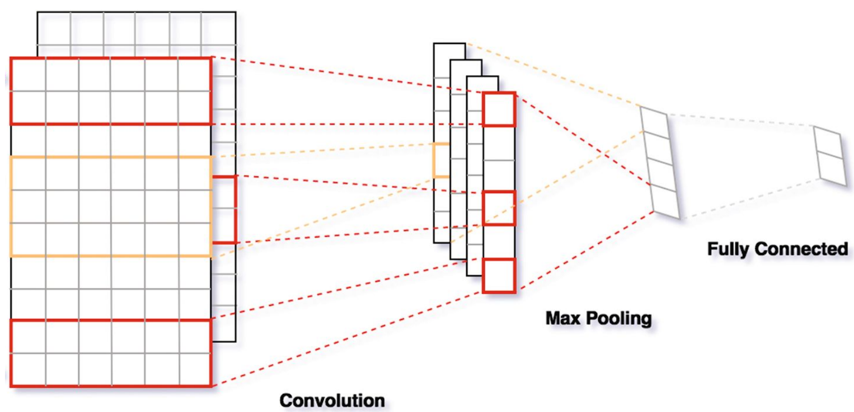


Figure 3: Standard CNN model including convolution, pooling, fully connected, and dropout layer. Adapted from [29].

Convolution Layer: this layer uses convolution operations to process information by surfing a fixed-size filter, or kernel, on the input data to obtain more refined data.
Pooling Layer: The main purpose of this layer is to perform pooling on the vectors outputted from the convolution layer, to allow only the most important vectors to be retained.
Fully Connected Layer: in CNN there always has one or several fully connected layer after convolution layer. This is just a typical perceptron-based technique to produce the final output, which is used to re-train again for the whole system in a back propagation manner.
Dropout: Technique used to prevent overfitting. During the training process, we use a probability p to randomly prevent some certain weights to be updated [29].

2.3 Long Short-Term Memory

It is a modification of the Recurrent Neural Network (RNN). Corresponding to data entry of input data with a linear pattern referred to as state, is the usual arrangement of RNN. State corresponds to a word for handling a text document. Input includes the corresponding data entry and output of the previous state. Output of each layer- weight matrix *W*.

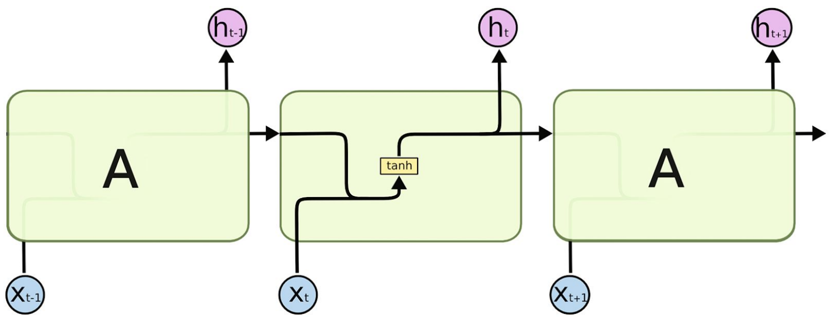
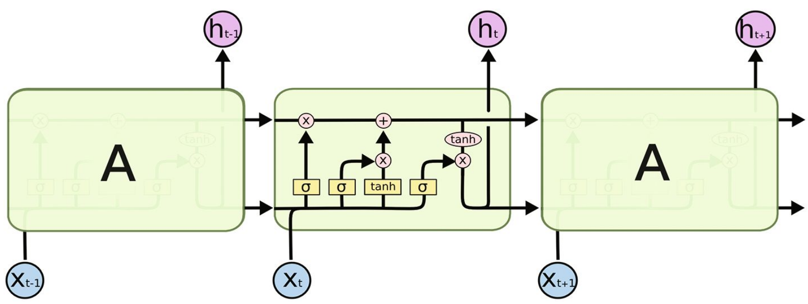


Figure 4: The standard model with a single roller RNNs layer [32]

Instead of a single layer neural network, LSTM has 3 gates with each layer having a *forget fate* deciding whether previously learned information isto be used for the current layer or not.

Figure 5: Parts of an LSTM cell [32]



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \;+\; b_f\right)$$

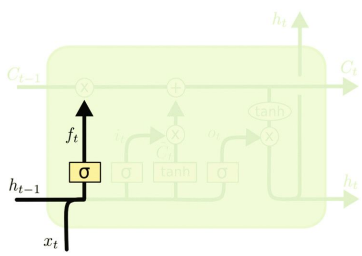


Figure 6: LSTM forget gate layer [29]

2.4 **Natural Language Processing basics**

Part-of-Speech Tagging is a process by which we assign a suitable part-of-speech to any word in the sentence. Preprocessing of headlines is done using-

- POS-Tagger:
  POS-Tagger or POST is used to find which word is which part of speech. NLTK library for python can be used to generate fairly good results)
- Lemmatization :
  Grouping different inflicted forms of words so they are analyzed as one. This task is done using mappings. Example: “is”, “are”, “have” are all converted to “be”. “go”, “went” are converted to “go”.
- Stemming:
  Stemming further reduces words to their root form. These are rule based which remove common prefixes and suffixes like “ing”, “ed” from words to convert them to a common base form.

2.5 **MultiLayer Perceptron**

Multilayer perceptron is useful for classifying data sets that are linearly separable. The Perceptron consists of an input layer and an output layer which are fully connected. MLPs have the same input and output layers but may have multiple hidden layers in between the aforementioned layers. MLPs form the basis for all neural networks and have greatly improved the power of computers when applied to classification and regression problems. Computers are no longer limited by XOR cases and can learn rich and complex models thanks to the multilayer perceptron.

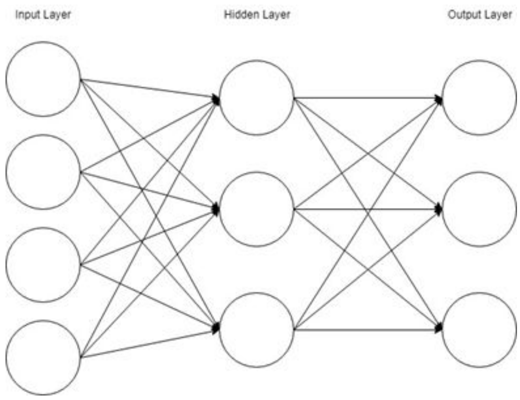


Figure 7: Architecture of MultiLayer Perceptron

2.6 **Sentiment Analysis**

Sentiment analysis is the interpretation and classification of polarity (positive, negative and neutral) within text data using text analysis techniques. Sentiment analysis, however, helps businesses discern unstructured text by automatically tagging it. With the recent advances in deep learning, the ability of algorithms to analyse text has improved considerably. Sentiment analysis uses various Natural Language Processing (NLP) methods and algorithms. The main types of algorithms used include:

A. Rule-based systems that perform sentiment analysis based on a set of manually crafted rules.
B. Automatic systems that rely on machine learning techniques to learn from data.
C. Hybrid systems that combine both rule-based and automatic approaches.

Rule-based Approaches

Rule-based systems use a set of human-crafted rules to help identify subjectivity, polarity, or the subject of an opinion.

Automatic Approaches

Automatic methods, contrary to rule-based systems, don't rely on manually crafted rules, but on machine learning techniques. A sentiment analysis task is usually modeled as a classification problem, whereby a classifier is fed a text and returns a category, e.g. positive, negative, or neutral.

The Training and Prediction Processes

In the training process (a), our model learns to associate a particular input (i.e. a text) to the corresponding output (tag) based on the test samples used for training. The feature extractor transfers the text input into a feature vector. Pairs of feature vectors and tags (e.g. positive, negative, or neutral) are fed into the machine learning algorithm to generate a model. In the prediction process (b), the feature extractor is used to transform unseen text inputs into feature vectors. These feature vectors are then fed into the model, which generates predicted tags (again, positive, negative, or neutral).

Feature Extraction from Text

The first step in a machine learning text classifier is to transform the text extraction or text vectorization, and the classical approach has been bag-of-words or bag-of-ngrams with their frequency. More recently,

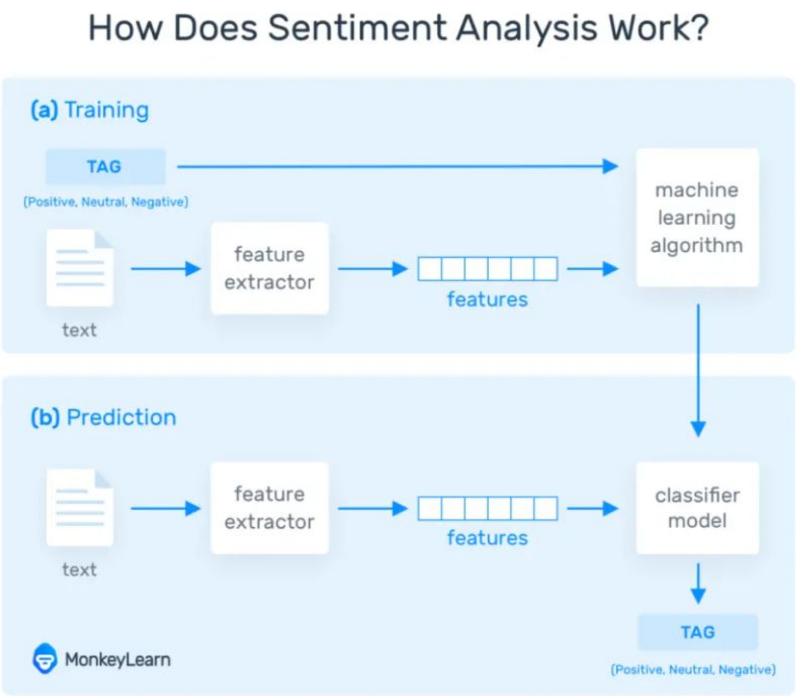


Figure 8: Working of automated sentiment analysis. Adapted from https://monkeylearn.com/sentiment-analysis/

new feature extraction techniques have been applied based on word embeddings (also known as word vectors). This kind of representation makes it possible for words with similar meaning to have a similar representation, which can improve the performance of classifiers. Common classification algorithms used are Naïve Bayes, Linear Regression, Support Vector Machines, Deep Learning. For this study we employ BERT embeddings to get the underlying sentence vector.

2.7 Word Embedding

Traditional word level embeddings like Word2vec, Glove, FastText, etc. incorporates possible meanings of words derived from surrounding sentence structure. Paragraph texts consist of multiple sentences which are coherent units of natural languages that convey information at a pragmatic or discourse level. Word embeddings give unsatisfactory results in a scenario where the global context differs from the local sentence structure.

- Basic Model: is used to encode the word in a dictionary of M words into a vector of length M usually as one-hot vector. Calculation of similarity and difference between two vectors result in the same value in Euclidean space and hence doesn't provide coveted accuracy.
- Word2vec: Word2vec represents the form of a distribution relationship of a word in a dictionary with the rest (known as Distributed Representation) [29]. Each element of the vector is characterized by a new learning value that is learned/updated by the Neural Network.

2.8 Belt and Road Initiative

According to a report by Asian Development Bank, developing Asia will need to invest $26 trillion from 2016 to 2030, or $1.7 trillion per year, if the region is to maintain its growth momentum, eradicate poverty, and respond to climate change (climate-adjusted estimate). Without climate change mitigation and adaptation costs, $22.6 trillion will be needed, or $1.5 trillion per year (baseline estimate) [46]. The

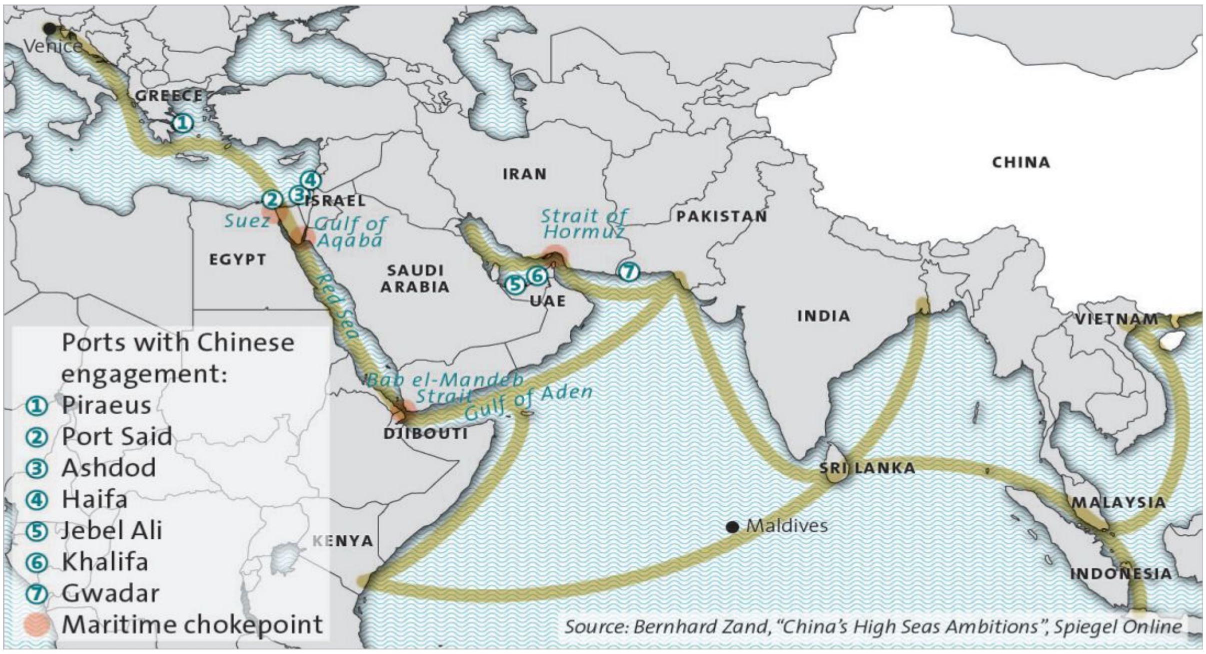


Figure 9: 21st Century Maritime Silk Road of BRI. Adapted from Spiegel International

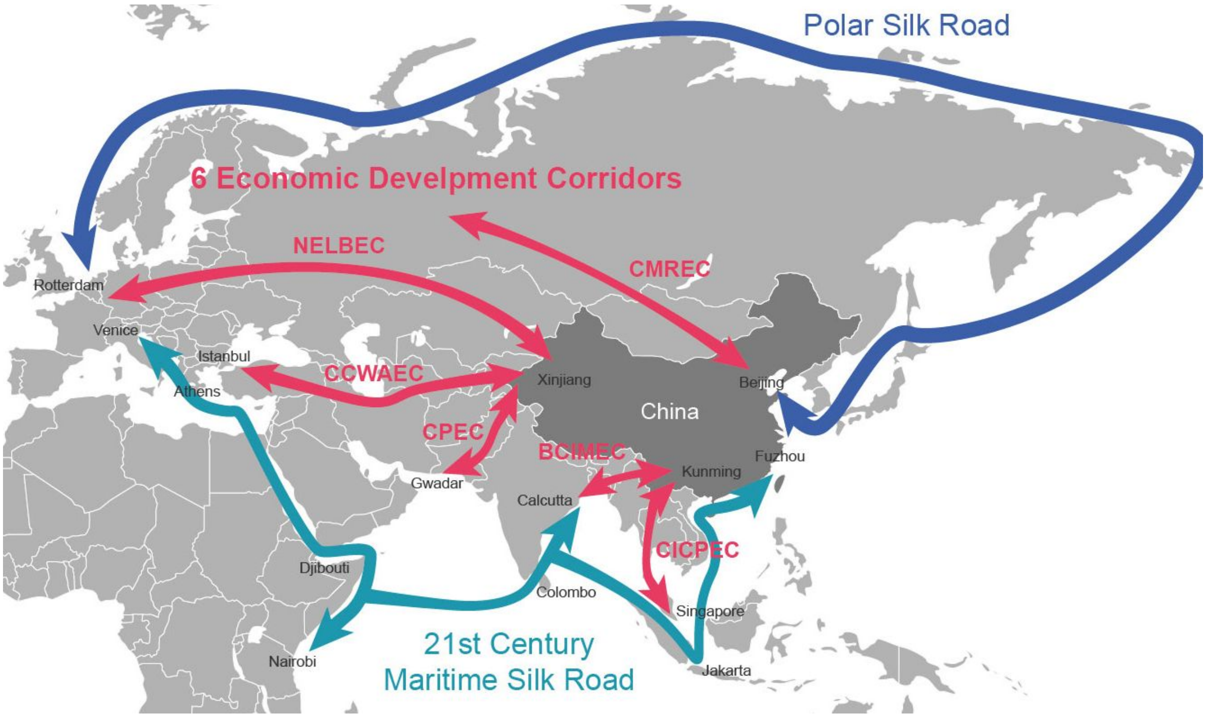


Figure 10: Silk Road Economic Belt and 6 Economic corridors. Adapted from www.beltroad-initiative.com

the report also concludes- "The infrastructure investment gap—the difference between investment needs and current investment levels—equals 2.4% of projected GDP for the 5-year period from 2016 to 2020 when incorporating climate mitigation and adaptation costs. Without the People's Republic of China (PRC), the gap for the remaining economies rises to a much higher 5% of their projected GDP. Fiscal reforms could generate additional revenues equivalent to 2% of GDP to bridge around 40% of the gap for these economies. For the private sector to fill the remaining 60% of the gap, or 3% of GDP, it would have to increase investments from about $63 billion today to as high as $250 billion a year over 2016–2020".

'Belt and Road' as communicated by the Chinese government is a concept which aims to increase connectivity between the Asian, European and African continents. The intention is for this increased connectivity to enhance trade flows and spur long-term regional economic growth and development, benefiting all those involved. In actuality, BRI is an 'umbrella' initiative. It seems to be a potentially huge collective of current, planned and future infrastructure projects, accompanied by a host of bilateral and regional trade agreements. Ongoing and planned projects will focus on the development of a wide array of assets, including ports, roads, railways, airports, power plants, oil and gas pipelines and refineries, and Free Trade Zones, etc., as well as supporting IT, telecom and financial infrastructure [51]. It seeemingly encomposses all FDI flowing out of China. Since its inception in 2013, BRI is valued at $3.87+ trillion already [52]. BRI is part of China's "going global" strategy. The initiative will take around 30-40 years to finish and by the end will put China at the center of international trade.

China was able to undertake the initiative as, while the biggest world economies were severely affected by 2008-09 financial crises, China was relatively unaffected and by 2012, it held a huge forign cash reserve of $3.5+ Trillion which was partly due to high volume of exports from Chinese market.

BRI can be regarded as the most important driver for China's long-term development strategy, foreign policy and – in the opinion of some – even a way to reform the structure of its own economy. China's enduring emphasis on heavy industries over the past two decades, as well as the government being a decisive force in the country's economy are two of the key reasons for this overcapacity.

BRI consists of land corridors-

- The New Eurasian Land Bridge, which runs from Western China to Western Russia through Kazakhstan, and includes the Silk Road Railway through China's Xinjiang Autonomous Region, Kazakhstan, Russia, Belarus, Poland and Germany.
- The China–Mongolia–Russia Corridor, which will run from Northern China to the Russian Far East.
- The China–Central Asia–West Asia Corridor, which will run from Western China to Turkey.
- The China–Indochina Peninsula Corridor, which will run from Southern China to Singapore.
- The China–Pakistan Economic Corridor (CPEC)

The land corridor Bangladesh–China–India–Myanmar (BCIM) Economic Corridor, since the 2nd Belt and Road Forum in 2019, BCIM has been **dropped** from the list of covered projects due to India's refusal to participate in the Belt and Road Initiative.

Key opportunities-

- Outbound capital projects and infrastructure – especially in partnership with Chinese players
- Supply equipment/technology/intellectual property
- Joint or independent engineering, procurement and construction/project finance
- Joint new client developments (e.g. developing market governments)
- Leverage Chinese partnerships abroad for accessing Chinese market itself
- Leverage Chinese funding for divestment, fundraising, etc.
- Outbound financing/private equity fund (e.g. joint AIIB, Silk Road Fund, etc.)
- Better trade with markets with improved infrastructure (Europe eastward)

It has been dubbed as the 'Chinese Marshall Plan'. However, the comparison is not an accurate one, in the sense that China is putting a very clear emphasis on the inclusiveness and 'win-win' character of its B&R initiative. Where the original Marshall plan deliberately excluded some countries from participation and put hefty conditions on others, China is making clear that all countries along the way are welcomed to join, without BRI attaching additional conditions [52-54].

This is one of the most important reasons as to why so many countries want to be part of the Belt and Road Initiative. China's Five Principles of Peaceful Coexistence and its non-interference and no pre-conditions pertaining to the ruling system of the country, makes the BRI more attractive than loans from established authorities like the World Bank and IMF[53]. List of banks in China financing the BRI is as follows-

Institution (excluding Ministry of Finance and Ministry Commerce Aid, etc.)	Features
China Development Bank (notably the world's largest development finance institution)	Non-concessional loans and credit lines Concessionary loans Overseas investment support Can be tied to exports in most cases Imposes limits to sovereign borrowers (such as the IMF) Controls concentration of loans Government capital injections and access to PBoC pledged Supplementary lending programme keeps funding very cheap
China Exim Bank	Preferential export credits (tied to exports) Export buyer's credit (tied to exports) Export seller's credit (tied to exports) Concessional loans (at least 50% are tied to exports) Non-concessional loans and credit lines (can be tied) Overseas investment support (can be tied) Debt ceilings for each country Government capital injections and access to the PBoC pledged Supplementary lending programme keeps funding very cheap
Agricultural Development Bank of China	Overseas investment support (can be tied to exports)
Industrial and Commercial Bank of China	Non-concessionary loans
Bank of China	Non-concessionary loans
Silk Road Fund	All BRI-related projects
China Construction Bank	Contributing to BRI related projects
New Development Bank (NDB)	To play a larger role in BRI projects
China Export and Credit Insurance Corporation	
Asia Infrastructure Investment Bank (AIIB)	Not BRI-related projects (China 36% voting)

Figure 11: Belt and Road Initiative Financing. Adapted from [54].

Chapter 3 **Literature Review**

In this chapter, we will discuss existing solutions for different phases of the project. We will also elaborate on our choice of model and model-specific components.

3.1 News Extraction and Crawler

To curate and compile data for a specific project, dismally, there are lack of publicly available integrated crawlers and extractor of news. Gathering news typically entails 2 phases-

- Crawling news websites
- Extracting information from news articles

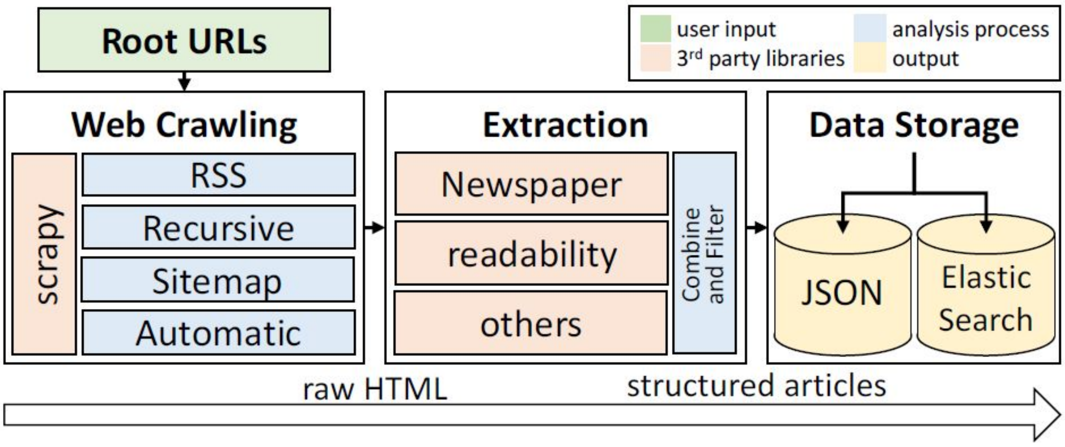


Figure 12: Architecture of *news-please* python api. Image taken from [10]

Crawling can be achieved by widely available web scraping frameworks such as *scrapy* for Python [7]. These types of crawlers are needed to be tailored for specific use-cases.
Extraction of information from news articles converts raw data received from the crawler into suitable format for analysis, such as Natural Language Processing. Website-specific extractors such as *yanytapi* [8] for New York Times articles, achieve high precision but require significant initial set-up for customization for specific needs. These are suitable for high-quality data with fewer websites to be crawled. Typical information to be extracted is headline, author's name, publication date, main text. Generic extractors like *boilerpipe*, *Goose*, *readability* render low performance than *Newspaper* [9]. It features robust extraction of major news article elements, supports 38+ languages but lacks full website extraction and news content verification. *Newspaper* lacks website extraction, auto-extraction of news articles, and news content verification.

In our project we employ *news-please* [10], a tool covering both crawling and extraction phase of news data and outperforming existing news collection python apis. The architecture of *news-please* is described in Figure 12. *news-please* covers 5 requirements-

- Broad coverage
- Full website extraction
- High quality of extracted information
- Ease of use, simple initial configuration

- Maintainability

The processes in Figure 12 are elaborated below-
<u>Web Crawling:</u>

- Crawler is downloaded using *scrapy*. Remainder of processes can be done in parallel.
- RSS: analyses RSS feeds from recent articles
- Recursive: follows internal links in crawled pages
- Site Map: analysing sitemap for link to all articles
- Automatic: tries sitemap, falls to recursive if error

<u>Extraction</u>

- *Newspaper, readability* along with regex parser is used for collection of author, headline, content, publish date, and leading paragraph of an article
- Combination of extractor is achieved using rule-based heuristics

3.2 News aggregators

Remarkably high percentage of people consult a small number of news source, despite exploding number of news outlets with little or no cast over the past decade, can cause narrow news perspective [11], and thus a skewed or incomplete perception of information. Media bias can significantly change a person's awareness and perception of a topic. This change can become critical for issues with high social, economic, and policy impact, such as elections [12,13]. A recent case being biased news coverage by prominent news media outlets is of Bernie Sanders campaign for Democratic Presidential Nomination 2020 [14-19].

Figure 13 depicts forms of media bias that could materialize in the news production process, in which publishers transform an absolute event into a news story [21]. During gathering, journalists select events, sources, and the facts they want to present. These selection processes bias the resulting news story. During writing, journalists can alter the reader's perception of a topic, e.g., through word choice (whether the author uses a positive or negative word to refer to an entity, such as "coalition forces" vs. "invasion forces"), or by varying the credibility ascribed to the source [23-24]. Third stage, editing, determines the (visual) presentation of an article, e.g., through placement (a front-page article receives the most attention). Finally, consumers read the news. Reading may also yield different perceptions of the event [25-27]. Existing news aggregators in literature consider the commonality between articles instead highlighting the differences between two articles. Other approaches to reduce media biases by broadcasting user's comprehension of news topics suffer from practical implementation, such as being confined by analysis of one news category[20,21], requiring manually created knowledge base [22], and being pre-tuned for specific tasks.

A recent approach is Matrix-based News Analysis (MNA) introduced in [27] attempts to tackle the issue of media bias. The authors use a 5 step approach for presentation and summarization-

A. Data gathering (crawling)
B. Article Extraction (from raw website data using website specific wrapper and content heuristic)
C. Grouping (finding groups related article about same topic)
D. Summarization
E. Visualization

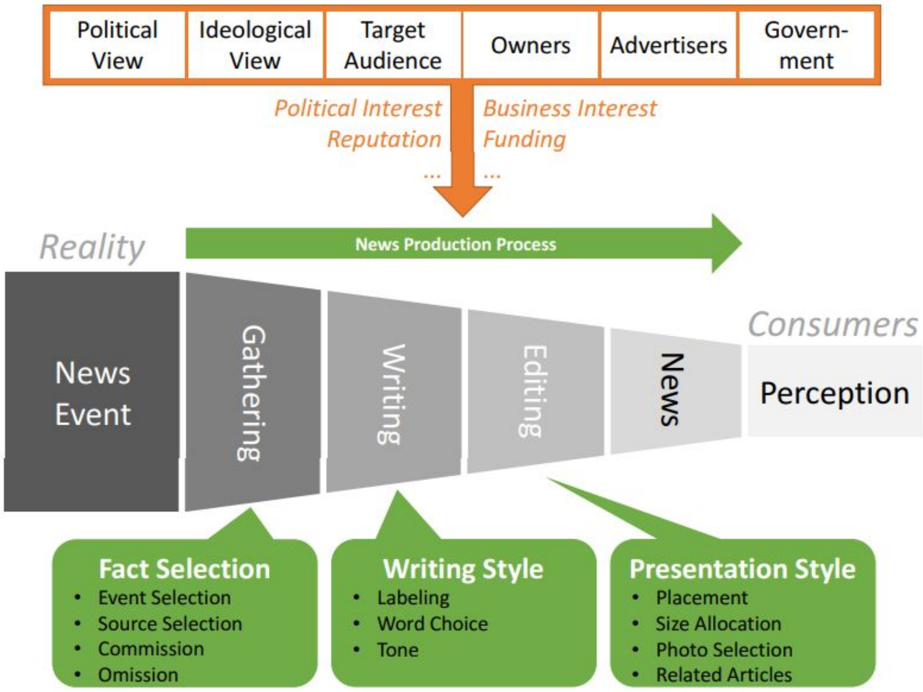


Figure 13: Reasons and forms of Media Bias. Image taken from [27]

Data extraction and article extraction are done using existing web-crawling frameworks. Grouping is done using Latent Dirichlet Allocation (LDA), which is also used in European Media Monitor, and hierarchical agglomerative clustering (HAC). Summarization is done using TF-IDF scores, redundancy, and order of appearance such as *MEAD.* Before Grouping step, additional steps are performed during article extraction which results in novelty-

- Text extraction (LDA performs better on long text)
- Topic Modelling (Parameter consideration of LDA is important in this step, 1500 min, iterations. Dirichlet hyperparameters as $\alpha=\beta=0.0001$.
- Post-processing (for topics and mappings with each cell has weighted proportion of each topic and removing all improprobable topic with weight<0.2 from a cell)
- Grouping documents (there is a trade off between precision and recall)

The document also explores using the sum of TF-IDF and topic weights to form the summary score. There are few assumptions taken that affect the scoring method drastically: it only considers the first 10 sentences of an article by the assumption that the first 10 sentences convey the sentiment of the article which is a gross mischaracterization as most articles consist of multiple subjects with varying opinion and analysis throughout. Nonetheless, it provides the motivation and direction of analysis that will be used in our analysis and experimentation.

The MNA model implementation in *NewsBird* provides a good alternative for media bias but also is ridden with fallacies. The reader is assumed to have a priori about the topic to be searched. Also it requires additional effort from the reader to scrutinize multiple aspects.

3.3 **Sentiment Analysis and Opinion Analysis**

Sentiment Analysis or Opinion Analysis aims at inferring sentiment orientation of a document. There are 3 levels of sentiment-

- Document-based level
- Sentence-based level
- Aspect-based level

In document and sentence -based analysis, it is implicitly assumed that the analyzed document/sentence only discusses a single object. Various methods have been put forward to understand the problem. Convolution-based techniques continue to be developed for sentence-based sentiment analysis [28]. Sentiment analysis of news articles differs from analysis of blogpost or movie reviews which show a clear sentiment and bias. News articles are objectively meant to be unbiased for journalistic purposes. Newspaper articles usually deal with multiple subjects interlinked together and mentioned multiple times. The length of a news article is also generally long.

3.3.1 Neural Network Approach

A popular approach to overcome these challenges, an architecture performing deep a document-level sentiment analysis for news articles can be considered [29]. To capture relationships made by the order of features, Recurrent Neural Network (RNN) systems such as Long Short-Term Memory have been used in combination with convolution processing to sentiment analysis for short text. The following is detailed working of state-of-the-art architecture in [29]-

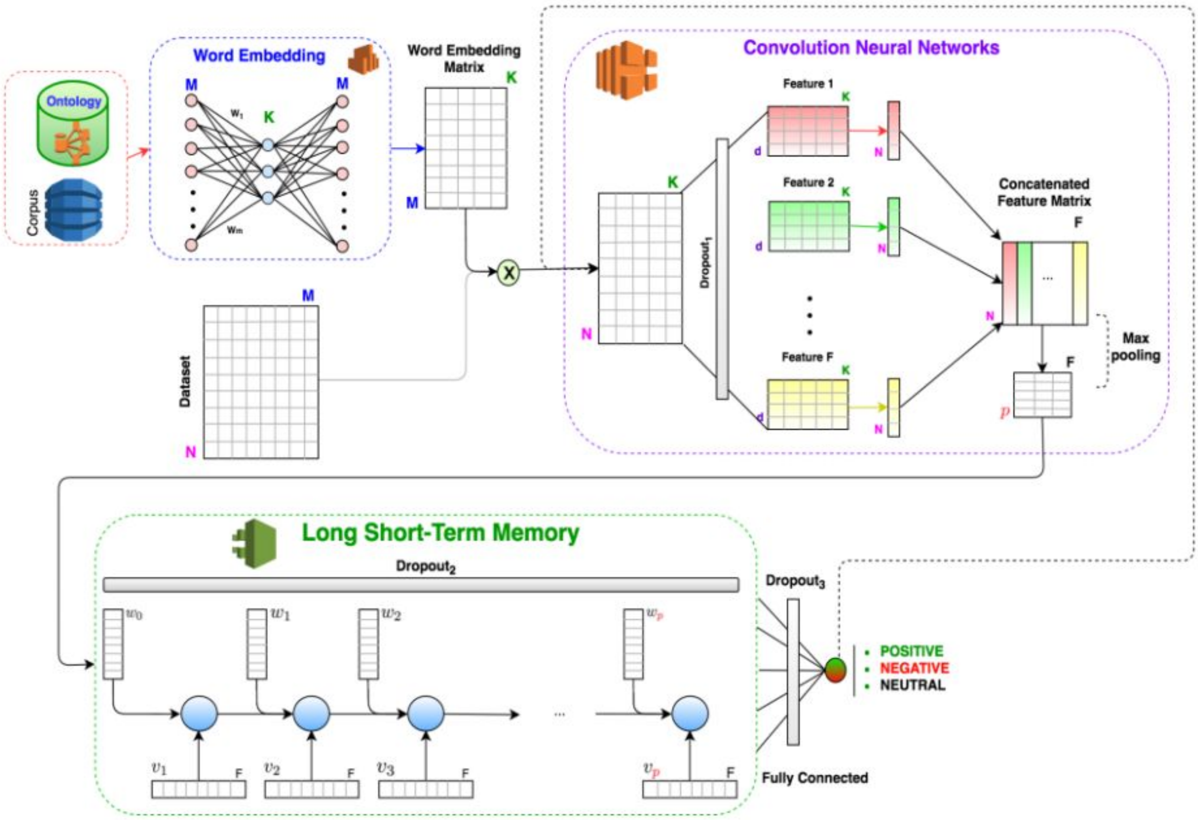


Figure 14: Deep Architecture for Sentiment Analysis. Adapted from [29]

Word Embedding(WE) Module: Comprises of 3 NN WE module-

- 1st layer: M inputs neurals where M= | words belonging to dictionary|
- Hidden Layer: K nodes (K <<M)
- 3rd Later: M nodes, back propagation to 1st layer
-

Training dataset: Set of collected news articles which are already labeled. Matrix from a document of N words of dimension NxM is created with row being one-hot vectors. Dimension vector and weight vector are multiplied to give the Embedding matrix of dimension NxK. It is used as input for the next CNN module. Uses SMOTE balancing, combines LSTM and convolution neural networks in an efficient way leading to promising results.

The only drawback is creation of the model is complex which requires careful parameter adjustment which is cumbersome. Additionally the dataset needs to be adequately large to provide good results. It can be combined with [45] to improve upon the results.

3.3.2 News Headline Analysis

Another approach used to analyze sentiment in a news article is through headlines. The rationale behind the approach is that most people judge the news contents directly by scanning news headlines only rather than going through the complete story and hence they generate a quick thought about it. This shows that even a small headline can also play a vital role in any judgment. Hence sentiment of the article along with summary of the context is represented in a concise form in the news headline and is bound to convey sentiments more strongly than the whole article [33].

Apoorv et al. [36] used SentiWordNet instead of naive-bayes classifier for sentiment analysis of news headlines as done in [34]. Authors of [36] use a Part-of-Speech Tagger and globally available resources of SentiWordNet. SentiWordNet is an opinion lexicon derived from the WordNet database where each term is associated with some numerical scores indicating positive and negative sentiment information [35].

3.4 **Relationship between news and public opinion**

In traditional agenda setting, “the core hypothesis is that the degree of emphasis placed on issues in the news influences the priority accorded these issues by the public” [38]. Examining the tone of news coverage, rather than simply the amount of coverage, is part of what is considered “second-level” or “attribute” agenda setting. This second level analyzes the attributes afforded issues and individuals in news coverage, whereas traditional agenda-setting research has focused primarily on amount and placement of news coverage [39]. In the second-level agenda setting, however, the hypothesis is that both the selection of topics for attention and the selection of attributes for thinking about these topics play powerful agenda-setting roles. It suggests, the media may go further than telling people what to think about; they may actually tell people how to think about a subject. This theory examined by Joe Bob Hester and Rhonda Gibson is the key inspiration of this project.

Scheufel [40] suggested that framing is part of agenda-setting approach, including second-level agenda setting, is based theoretically on the premise of attitude accessibility. It conveys the fact that the media has the power to portray a relatively neutral story in light they deem worthy to increase the salience of an issue.

In the paper [41] an argument the role of media is given- "*traditional surveillance function of news dictates that the media mirror public opinion and monitor social change, not purposefully influence such phenomena*". Which in turn implies that the masses tend to demand more awareness about incoming threats and hard, giving the media a negative outlook of conditions inherently. The study suggested that- "Their findings indicated that close to half of the public reported typically not getting any economic news. More important, news exposure did not lead to a significant improvement of ability to assess economic situations...The economy, unlike areas such as foreign affairs and national politics, is one with which people have extensive day-to-day personal experience". This leads to the notion that policy has deeper impact than news about economics.

3.5 **BERT**

Bidirectional Encoder Representations from Transformer (BERT) is a masked language model which encoder decoder network of transformers for bidirectional learning[11]. Additionally, it can be easily fine-tuned for any downstream task with few additional layers. BERT and its variants have outperformed traditional language models in a variety of tasks such as question answering and language inference[12, 13]. It can accurately learn the distribution of out of vocabulary words without substantial modifications [14]. Figure 15 shows overview of BERT architecture. BERT's key technical innovation is applying the bidirectional training of Transformer, a popular attention model, to language modelling. This is in contrast to previous efforts which looked at a text sequence either from left to right or combined left-to-right and right-to-left training. The paper's [49] results show that a language model which is bidirectionally trained can have a deeper sense of language context and flow than single-direction language models. When training language models, there is a challenge of defining a prediction goal. Many models predict the next word in a sequence, a directional approach which inherently limits context learning. To overcome this challenge, BERT uses two training strategies: Masked LM and Next Sentence Prediction. Thorough discussion and analysis of BERT is out of the scope of this project.

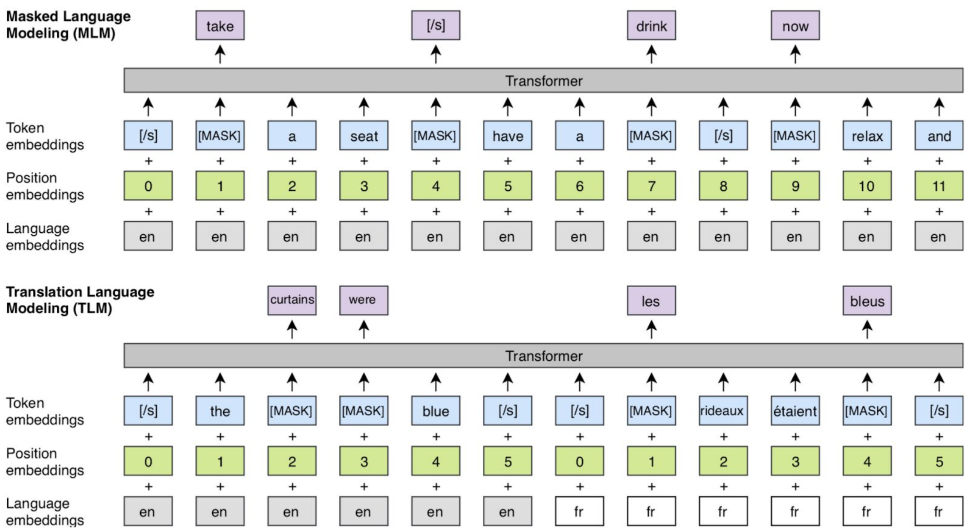


Figure 15: BERT's architecture

3.5 Belt and Road Initiative

Belt and road initiative promises a myriad and an unfathomable portfolio consisting of train network from China to London, gas pipeline from Caspian Sea to China, high speed railway in south-east Asia to Nuclear power plant and hydroelectric dams. Beijing portrays it as a cross-border, win–win economic stimulus package that will spur economic growth in China and also the countries with which it engages along the old Silk Roads [56]. China has pledged to finance and build infrastructure, creating new economic corridors that stretch across central Asia to Europe and south and south-east to the Indo-Pacific.

Through the Belt and Road Initiative, for many less developed regions, mostly in inland China, it is a clear opportunity to catch up with the more advanced provinces on China's East coast. Central government also intends to bring more stability to the interior states (most notably Xinjiang) by establishing better connectivity with other regions. Xinjiang border has witnessed growth of multiple cities and trading centers in the last 5 years. It creates a stable external environment for China's domestic economic growth. As a result, participation from countries in Southeast Asia is pivotal to the success or failure of Beijing's ambitious undertaking [57].

BRI mainly connects Eurasia to China through both "Belts" and maritime "Roads". The new Silk road expands to polar regions and Latin America as well and is much more comprehensive and thorough. On successful completion of BRI, China would have established itself not as the global trade powerhouse but also as go-to investment and loaning cog as well.

BRI also provides China with an outlet for its excessive production of cement and steel. China produces more than half of the total supply of cement (587% more than the second largest producer, India) and in 2017, China produced 1.1 Billion tonnes of steel which is as much as the rest of the world combined. Infrastructure projects backed by State-Owned-Enterprises construction behemoths would not only generate revenue but will allow Chinese companies to venture into previously untapped markets when growth-rate is slowing down at home.

BRI also employs China's workforce overseas, bridging cultural gaps too. A major commodity in the establishment of the Silk Road is oil. China has become the largest consumer of oil. The United States' shale oil boom in the past decade wasn't mirrored in China. China holds the world's largest reserves of shale natural gas, according to the US Energy Information Administration. However, much of that gas is considered recoverable only if cost were not a constraint. China's three biggest gas basins, the Ordos in northern China, the Tarim in the Xinjiang region in the west, and Sichuan province in the southwest, make up 90 per cent of the country's output, but each of them are beset with geological or technological difficulties. Drilling is more dangerous and expensive compared. The drilling also requires 2-3 times deeper than normal hydraulic fracturing wells. These shale deposits are in desert and some above fault lines.

To meet its enormous oil demand, a network of oil pipelines converging to China are being built. One of the most prominent is from Gwadar, Pakistan to West China. Oil tankers anchoring at Gwadar shaves additional time needed to go through Malacca Strait. This also creates an alternate route to avoid maritime chokepoints and save billions of dollars transportation costs. In some countries, like Venezuela, the majority of ($36 Billion 2013) loan repayment is through crude oil.

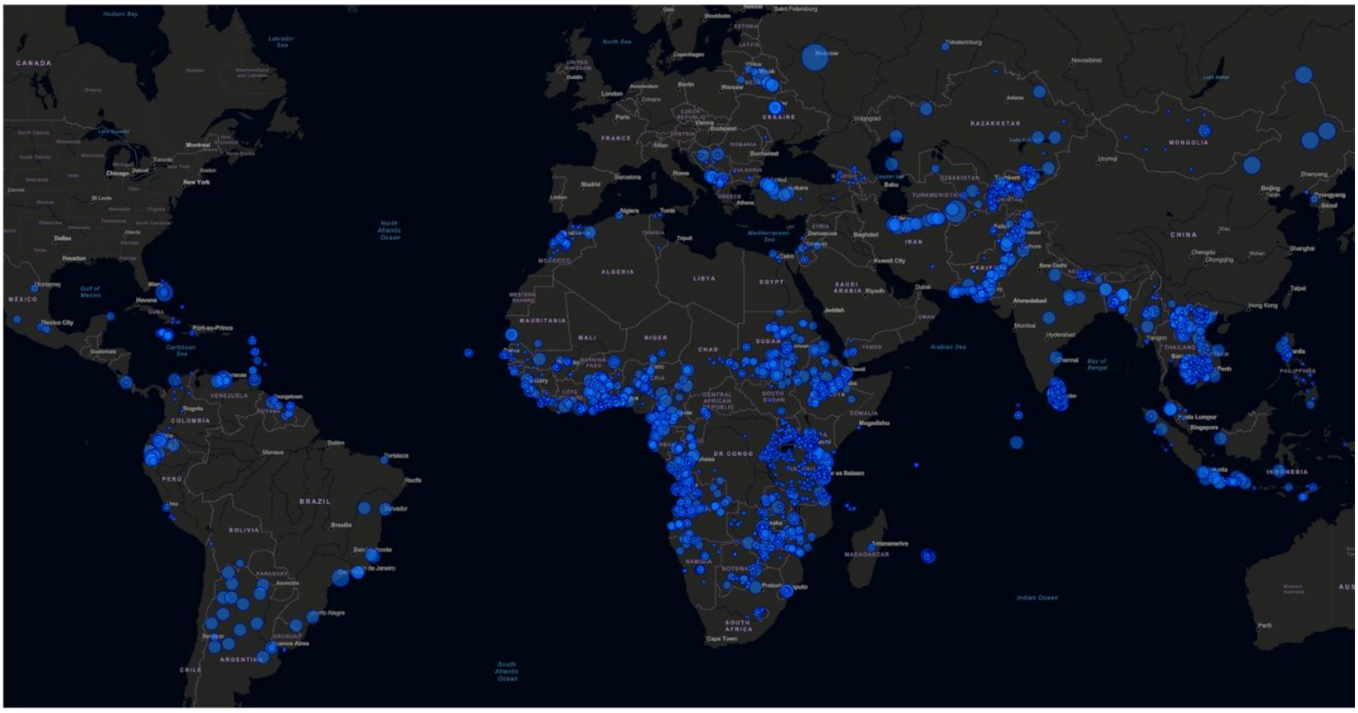


Figure 16: The circles on the map pinpoint the location of thousands of Chinese-funded development projects. The bigger the circle, the bigger the investment. The largest circles represent projects in the multibillion-dollar range. Map by Soren Patterson, AidData/William & Mary

Chinese loans and investments during the BRI has helped economies recover. In Ethiopia, $1 Billion loan for construction of Grand Renaissance Dam. Investments in Georgia has led to creation of the Tbilisi housing project, school, medical clinic, free-trade pact, mining project, Oil pipeline from Azerbaijan. Another lauded project is the railroad from Addis Adaba (capital of Ethiopia) to Djibouti. Providing $147 Million in food aid to Africa.

That shroud of secrecy has bolstered a popular narrative that China is a "rogue donor" that pours money into undemocratic governments to promote Chinese growth and access to natural resources. Skeptics say the Chinese projects are of little use to the countries, often in Africa, where they're built. Politicians have described some of them as "white elephant" projects. They point to seaports, bridges and other projects that cost a great deal to build but aren't actually getting a lot of use. Chinese financed “connective infrastructure” such as roads, bridges, railways and ports are distributing economic growth into rural areas more evenly than traditional Western development programs have. Observing nighttime satellite images from the U.S. Defense Meteorological Satellite Program's Operational Linescan System satellites. It's "reasonably well-established in [academic] literature" that nighttime light is a good indicator of household income – more light means that families in that area have more money [59]. The AidData researchers measured changes over time, starting back in 2000 and going up to 2013, in the amount of light visible within a certain distance from China's connective infrastructure. They found that in the later images, light was not just concentrated in the immediate vicinity of the projects, but it also had spread within the provinces and districts where they were built, as well as between provinces and districts. This, they say, suggests that Chinese connective infrastructure is spreading economic growth across large regions [58]. The main reasons why countries are gravitating towards China is due to weakening of world economic order and failure in expectations being met from institutions like the

World Bank. Crumbling of Trans-Pacific Partnership also resulted in erosion of stability associated with the United States.

Analysts also say that exclusion of China from Trans-Pacific Partnership and announcement of the “New Silk Road” by the USA for rebuilding Afghanistan in 2011 created the motivation for China to pursue BRI [61]. United States attempt for the “New Silk Road” was regarded as a disaster as the funds were too thinly spread and the developments were under strict regulations and governance as deemed by the USA. USA threatening to leave NAFTA (North Atlantic Free Trade Agreement) and castigating NATO alliances has led to a climate of mistrust and protectionism compared to dirigiste followed by China.

Countries have accused China as a neocolnist power. It is blamed for creating “white elephant” projects in debt-ridden countries and practices to date for infrastructure financing, which often entail lending to sovereign borrowers, then BRI raises the risk of debt distress in some borrower countries. “Debt-trap diplomacy” is a usual narrative mentioned in the news media [63].

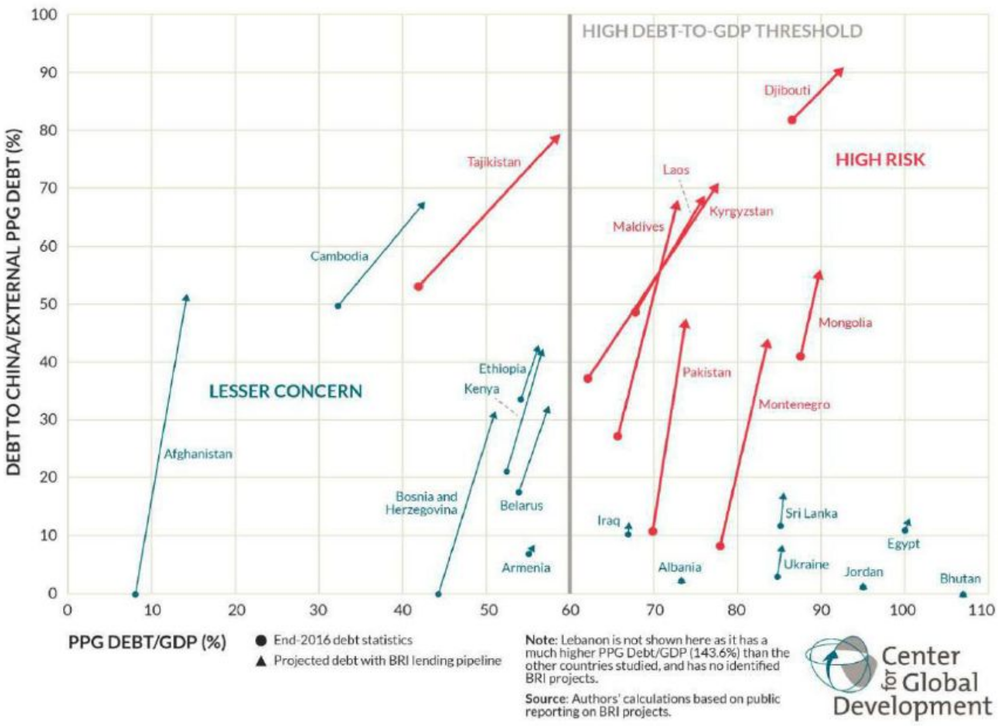


Figure 17: Immediate marginal impact of BRI lending pipeline. Adapted from [60].

Recent investments by Chinese companies in ports in Greece and elsewhere in Southern Europe have been called Trojan horses entering Europe through its *soft underbelly* [64-66]. China, by buying up ports, is said to be trying stealthily to expand its military presence along ancient trading routes and vital shipping lines. So far there is plenty of speculation, but no incontrovertible evidence of Chinese military strategy connected to the BRI.

On July 29, 2017, Sri Lanka sold a majority of shares in its loss-making Hambantota port to ChinaMerchants Port Holdings Co. for US$1.12 billion [67]. This transaction was characterized as an ‘asset seizure’ as though the Chinese had forcibly taken control of the port when the Sri Lankans were allegedly unable to repay the Chinese loans that had financed the port's construction. This resulted in Sri Lanka surrendering it's port to China for 99 year lease. The deal is associated with China trying to foster military leverage by acquiring a strategic position in Indian Ocean. This theory was widely regarded as true as in previous years, China had tried to acquire and claim islands in the South China Sea creating tension between various Pacific Islands and China [54, ,60, 62].

India, despite being part of Bangladesh–China–India–Myanmar (BCIM) Economic Corridor, declined in participating in BRI due to political reasons which will be explored in Section 6.

Chapter 4 **Data collection and condensation**

The data required was acquired in 3 stages.

- The first stage was url collection of relevant articles. It was done using CLI mode in *news-please*. The search phrase used for “belt and road China”. Using simply “belt and road” resulted in articles involving only the keyword: “road’ such as articles about road safety and road accidents. Some websites had limited search capabilities (Example some newspapers in Pakistan and Myanmar), for those countries, corpus was built by extraction using “site:<news website> belt and road China”. Further challenges were encountered for Forbes articles which were behind strong paywalls. Headlines and publication dates could be extracted from Forbes articles directly via their url but text collection had to be done using request api.
- The second stage was getting corresponding data from url. news-please was used for providing headline, date of publication, context, and keywords (done using nltk in *news-please*).
- Third stage was manually filling out missing data.

The dataset includes 500 articles from news outlets across various countries. Some data preprocessing was performed to create columns- ‘countries’. The values are dictionary with key as country mentioned in article text and value being the number of times it is mentioned in the article.

```
In [28]: train_data["Date of Publication"] = dates
         train_data["Headlines"] = headlines
         train_data["Text"] = texts
         train_data["Keywords"] = keywords_l
         #train_data["List of countries"] = list_of_countries
```

```
In [31]: for url in tqdm(train_data["html link"]):
             dict_country = {}
             #url = "https://www.washingtonpost.com/opinions/this-crisis-has-taught-us-true-cost-of-doing-business-with-china/2020/04/20/
             if "forbes" in url:
                 list_of_countries.append("")
                 continue
             try:
                 a = Article(url)
                 a.download()
                 a.parse()
                 a.nlp()
             except:
                 list_of_countries.append("")
                 continue
             try:
                 for country in pycountry.countries:
                     if(country.name == "United States"):
                         country_count = a.text.count("US") + a.text.count("USA") + a.text.lower().count(country.name.lower())
                     else:
                         country_count = a.text.lower().count(country.name.lower())
                     if(country_count>0):
                         dict_country.update({country.name: country_count})
                 list_of_countries.append(dict_country)
             except:
                 list_of_countries.append("")

100%|████████████████████████████████████████████████████████████| 500/500 [07:15<00:00,  1.23s/it]
```

Figure 18: Code snippet from data_analysis.ipynb for obtaining country list column

The column Eastern/western was added via manually identifying which media sources can be considered as Western media. Countries belonging to NATO, EU, or Europewere considered western. Exceptionally, India was also considered under western media. The rationale being that India being opposed to BRI and CPEC followed rhetoric similar to the USA. The reason and analysis of this exception is explored further, later in the paper.

The most important column of the data is undoubtedly "score" which is the sentiment assigned to the article. The sentiment assigned by the author was under strict impartiality and indifference. For assigning score values, strict guidelines laid down under [25, 47]. The mentioned papers theorize that there are three different views of an article: author's, readers', and text which is represented in Figure 19.

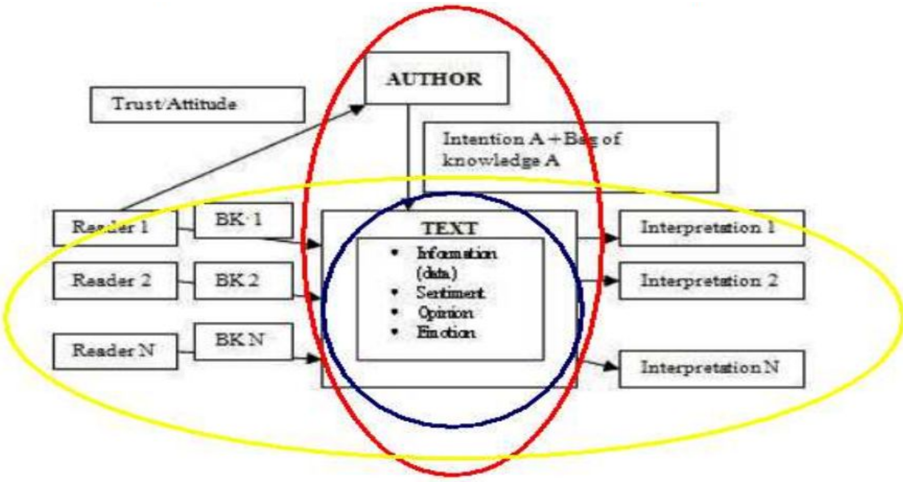


Figure 19: Three components of text opinion. Adapted from [25]

Authors in [47] further claim that opinion in an article can also be introduced via omission of facts. [48] observed- (a) quotations typically contain more sentiment expressions than other parts of news articles and (b) source and target of opinion are better defined than in document level opinion mining, although the second set of experiments has shown that entities mentioned in quotations are not necessarily the target of the sentiment. [47,48] both assert that opinion as a subjective identity which may not imply a sentiment. A negative news report impartially will carry a score of zero, proceeding through above guidelines.

Figure 30: Snapshot of first 4 data rows of dataset

	name of publication	html link	score	Date of Publication	Headlines	Context	Keywords	List of Countries
1	Washington Post	https://www.washington|	-1	2020-04-20 00:00:00	This crisis has taught us the true cost of d	Public health is not the only area in	['countries', 'global', 'debt', 'taught', 'banks', '	{'China': 18, 'Germany': 1, 'Sri Lanka': 1, 'Niger': 1, 'Nigeria': 1, 'Pakistan': 1, 'United States': 1, 'Zambia': 1}
2	Washington Post	https://www.washington|	-1	2019-04-23 00:00:00	The forgotten victims of China's Belt and F	But one group of victims is often ove	['initiative', 'forgotten', 'projects', 'policies', '	{'Bahamas': 1, 'Belarus': 1, 'China': 10, 'Ethiopia': 1, 'Israel': 1, 'Northern Mariana Islands': 1}
3	Washington Post	https://www.washington|	-1	2018-08-27 00:00:00	China's debt traps around the world are a	But the 93-year-old leader, who reca	['projects', 'port', 'debt', 'malaysia', 'chinese',	{'Australia': 1, 'China': 14, 'India': 2, 'Japan': 1, 'Cambodia': 1, 'Sri Lanka': 2, 'Myanmar': 1, 'Montenegro': 3, 'Malaysia': 7, 'Nepal': 1, 'Pakistan': 13, 'Serbia': 1}
4	Washington Post	https://www.washington|	-1	2018-08-22 00:00:00	Why countries might want out of China's I	"The future's coming now," a group o	['countries', 'projects', 'beijing', 'chinese', 'be	{'China': 15, 'Djibouti': 1, 'Kenya': 1, 'Kyrgyzstan': 1, 'Sri Lanka': 2, 'Maldives': 1, 'Myanmar': 1, 'Montenegro': 1, 'Mongolia': 1, 'Malaysia': 7, 'Pakistan': 2, 'Tajikistan': 1, 'Turkey': 1, 'United States': 1}

Using the above mentioned principles, the author of this paper adopted the following scoring mechanism-

A. Score of an article will belong to the set {-1,-0.7,-0.3,0,0.3,0.7,1}
B. Positive score implies that the article overall "praises" China for BRI.
C. Negative score implies the article is a negative critique of BRI.

D. Neutral score implies that the article is either reporting of facts or carried in-depth analysis of both sides of the story to create an impartial article.
E. -1: ["China induces debt-trap and is not reliable", "China's loan designed to steal assets"]
F. -0.7: ["BRI is full of malpractices", "BRI is harmful for the environment", "BRI project are riddled with corruption", "USA policies are more favorable than China's policies"]
G. -0.3: ["BRI is rooted in geopolitics", "BRI is ad alternative to free-state capitalism"]
H. 0.3:["China's intention is being misrepresented", "BRI project has potential for positive change to the community involved"]
I. 0.7: ["China's policies have created positive difference", "BRI projects and loans are better alternative than western counterparts"]
J. 1: ["China's BRI is being misrepresented as debt-trap diplomacy", "Praise for BRI and the projects:]

The sentiment is also given considering quotes mentioned in the article. For complete **unbiased opinion marking** of data, the article text was converted to text file with randomized file being opened at a time for marking the score making it impossible to tell which media outlet did the article belong to.

Chapter 5 Experimentation and Results

5.1 Sentiment analysis using proposed pre-processing

<u>Claim 1</u>: Rudimentary Anaphora resolution of article combined with headline represents the sentiment of article more prominently

In our experimentation we will use a pre-trained model for training and prediction processes from Google's luminary BERT (Bidirectional Encoder Representations from Transformers) [30] and Python Flair (library based on BERT and Pytorch) [31]. For classification we will use multilayer perceptron.

<u>Experiment 1:</u>

Three separate models were trained using the above configuration. Model accuracy was inconsistent and very low when *flair* library was used. The result was not considered for evaluation. The classification obtained through *flair* isn't discarded and is presented in the github repository of the project.

Three new separate models was trained using above configuration for performing sentiment analysis:
Model 1: Text for training and classification is complete text of an article.
Model 2: Text for training and classification only considers headlines of an article.
Model 3: Text for training and classification is *pre-processed and shortened* with headlines of an article.

Rationale for the experiment is as follows-
As considered in Section 3.3.2, headlines carry over sentiment of an article.
Pre-processing of the text is shortening of the article. The shortening is done by the following rules-

- As a news article contains multiple subjects and analysis, the sentiment associated with a particular subject can be obtained by extracting sentences with the target word as subject.
- Anaphora Resolution refers to extraction of text when the target word is used by an anaphora (the use of "it" or "that" to refer to the target word). A rudimentary resolution was performed by selecting 2 sentences before a particular sentence and 2 sentences after the sentence identified.
- Sentences of interests were selected not only based on target words like "debt", "BRI", "China", "Xi" etc. but also sentences with a county's name are deemed important and are resolved anaphorically.

These selected articles are then combined (with redundant sentences being deleted) to shorten the article and provide concise sentiment of the subject. The author believes such pre-processing will result in a better model. Anaphora resolution and headline analysis have always been considered separately in literature.

```python
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")

class BertForMultiLabelSequenceClassification(BertPreTrainedModel):
    """BERT model for classification.
    This module is composed of the BERT model with a linear layer on top of
    the pooled output.
    """
    def __init__(self, config, num_labels=7):
        super(BertForMultiLabelSequenceClassification, self).__init__(config)
        self.num_labels = num_labels
        self.bert = BertModel(config)
        self.dropout = torch.nn.Dropout(config.hidden_dropout_prob)
        self.classifier = torch.nn.Linear(config.hidden_size, num_labels)
        self.apply(self.init_bert_weights)

    def forward(self, input_ids, token_type_ids=None, attention_mask=None, labels=None):
        _, pooled_output = self.bert(input_ids, token_type_ids, attention_mask, output_all_encoded_layers=False)
        pooled_output = self.dropout(pooled_output)
        logits = self.classifier(pooled_output)

        weights = torch.from_numpy(np.array([0.15, 0.15, 0.15, 0.10, 0.15, 0.15, 0.15])).type(torch.cuda.FloatTensor)

        if labels is not None:
            loss_fct = CrossEntropyLoss(weight=weights)
            loss = loss_fct(logits.view(-1, self.num_labels), labels.type(torch.cuda.LongTensor).view(-1))
            return logits, loss
        else:
            return logits

    def freeze_bert_encoder(self):
        for param in self.bert.parameters():
            param.requires_grad = False

    def unfreeze_bert_encoder(self):
        for param in self.bert.parameters():
            param.requires_grad = True
```

Figure 20: Snippet of code for BERT for sentiment analysis in this project (./BERT/deepmodels.py)

As the model is pre-trained, we only need to train the last layers instead of the whole model (Optimization scope concluded from failed experiment). For obtaining better accuracy, the scores below zero (i.e -0.3 and -0.7) were approximated to -1 and scores above 0 (i.e. 0.3 and 0.7) were approximated to 1. Such approximations will give better results as the corpus isn't proportionate for each class (Optimization scope concluded from failed experiment).

Running BERT gives the output of a feature vector of length **768**. This feature vector is dumped into a file using pickle and then sent to the Multilayer Perceptron for classification.

```
**************original*****************
/home/shantanu/.local/lib/python3.6/site-packages/sklearn/neural_network/multilayer_perceptron.py:566: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (200) reached and the optimization hasn't converged yet.
  % self.max_iter, ConvergenceWarning)
{'-1': {'precision': 0.5555555555555556, 'recall': 0.35714285714285715, 'f1-score': 0.43478260869565216, 'support': 14}, '0': {'precision': 0.6666666666666666, 'recall': 0.7692307692307693, 'f1-score': 0.7142857142857142, 'support': 26}, '1': {'precision': 0.4444444444444444, 'recall': 0.5, 'f1-score': 0.47058823529411764, 'support': 8}, 'accuracy': 0.6041666666666666, 'macro avg': {'precision': 0.5555555555555556, 'recall': 0.5421245421245421, 'f1-score': 0.5398855194251614, 'support': 48}, 'weighted avg': {'precision': 0.5972222222222222, 'recall': 0.6041666666666666, 'f1-score': 0.59214772865668, 'support': 48}}
{'-1': {'precision': 0.6326530612244898, 'recall': 0.5, 'f1-score': 0.5585585585585586, 'support': 124}, '0': {'precision': 0.6741071428571429, 'recall': 0.755, 'f1-score': 0.7122641509433962, 'support': 200}, '1': {'precision': 0.5754716981132075, 'recall': 0.5865384615384616, 'f1-score': 0.5809523809523809, 'support': 104}, 'accuracy': 0.6401869158878505, 'macro avg': {'precision': 0.6274106340649467, 'recall': 0.6138461538461538, 'f1-score': 0.6172583634847786, 'support': 428}, 'weighted avg': {'precision': 0.6381295905771937, 'recall': 0.6401869158878505, 'f1-score': 0.6358250912826826, 'support': 428}}
**************heads*****************
/home/shantanu/.local/lib/python3.6/site-packages/sklearn/metrics/classification.py:1437: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples.
  'precision', 'predicted', average, warn_for)
{'-1': {'precision': 0.3333333333333333, 'recall': 0.07142857142857142, 'f1-score': 0.11764705882352941, 'support': 14}, '0': {'precision': 0.5333333333333333, 'recall': 0.9230769230769231, 'f1-score': 0.676056338028169, 'support': 26}, '1': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 8}, 'accuracy': 0.5208333333333334, 'macro avg': {'precision': 0.2888888888888889, 'recall': 0.3315018315018315, 'f1-score': 0.26456779895056615, 'support': 48}, 'weighted avg': {'precision': 0.38611111111111107, 'recall': 0.5208333333333334, 'f1-score': 0.40051090858878763, 'support': 48}}
{'-1': {'precision': 0.5769230769230769, 'recall': 0.24193548387096775, 'f1-score': 0.3409090909090909, 'support': 124}, '0': {'precision': 0.5138888888888888, 'recall': 0.925, 'f1-score': 0.6607142857142857, 'support': 200}, '1': {'precision': 0.6875, 'recall': 0.10576923076923077, 'f1-score': 0.18333333333333335, 'support': 104}, 'accuracy': 0.5280373831775701, 'macro avg': {'precision': 0.5927706552706552, 'recall': 0.4242349048800662, 'f1-score': 0.39498556998557, 'support': 428}, 'weighted avg': {'precision': 0.5743370077482227, 'recall': 0.5280373831775701, 'f1-score': 0.45206133430432494, 'support': 428}}
**************shortened*****************
{'-1': {'precision': 0.25, 'recall': 0.07142857142857142, 'f1-score': 0.11111111111111112, 'support': 14}, '0': {'precision': 0.5609756097560976, 'recall': 0.8846153846153846, 'f1-score': 0.6865671641791045, 'support': 26}, '1': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 8}, 'accuracy': 0.5, 'macro avg': {'precision': 0.2703252032520325, 'recall': 0.31868131868131866, 'f1-score': 0.2658927584300719, 'support': 48}, 'weighted avg': {'precision': 0.37677845528455284, 'recall': 0.5, 'f1-score': 0.40429795467108903, 'support': 48}}
{'-1': {'precision': 0.68, 'recall': 0.4112903225806452, 'f1-score': 0.5125628140703519, 'support': 124}, '0': {'precision': 0.5683890577507599, 'recall': 0.935, 'f1-score': 0.7069943289224953, 'support': 200}, '1': {'precision': 0.6666666666666666, 'recall': 0.15384615384615385, 'f1-score': 0.25, 'support': 104}, 'accuracy': 0.5934579439252337, 'macro avg': {'precision': 0.6383519081391422, 'recall': 0.5000454921422665, 'f1-score': 0.4898523809976157, 'support': 428}, 'weighted avg': {'precision': 0.624605478699732, 'recall': 0.5934579439252337, 'f1-score': 0.539618352171081, 'support': 428}}
```

Figure 21: Results obtained from sentiment analysis (./BERT/results.txt)

Results 1:

The result of experiment 1 as follows. These figures are averaged across five folds.-

- Model 1 (with full text) gave an accuracy of 0.5
- Model 2 (with only headlines) gave an accuracy of 0.52
- Model 3 (combination of shortened data + headlines) gave an accuracy of 0.6

Hence, claim 1 is **correct**.

5.2 Data analysis of news database for BRI

Claim 2A: Opinion of BRI is biased "against" leaning for western media and "pro" leaning for eastern media

Experiment 2A:

Addition of all scores and averaging of scores can give evaluation of bias, if any. The results are not expected to have 'high' modular value. Most articles are inherently non-biased and results of average should be closer to 0 rather than closer to 1.

Result 2A:

```
In [8]: new_train = pd.read_csv("data/China.csv")
```

```
In [9]: new_train.groupby('eastern/western')['score'].mean().plot.bar()
        print(new_train.groupby('eastern/western')['score'].mean())
        new_train.groupby('eastern/western')['score'].sum().plot.bar()
        print(new_train.groupby('eastern/western')['score'].sum())
```

```
eastern/western
E    0.132222
W   -0.111178
Name: score, dtype: float64
eastern/western
E    23.8
W   -36.8
Name: score, dtype: float64
```

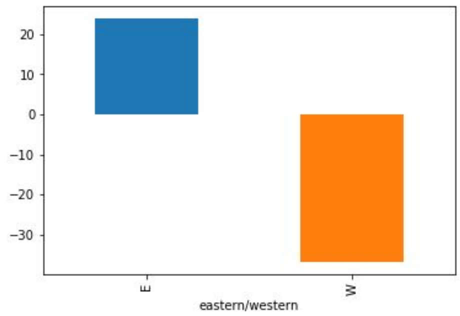


Figure 22: Result obtained from Experiment2A (./csv_analysis.ipynb)

Average score obtained for western media is **-0.1112**. Negative score implies bias against BRI.
Average score obtained for eastern media is +**0.1322**. Positive score implies bias for BRI.
Hence Claim 2A is **correct.**

<u>Claim 2B</u>: Negative media coverage of BRI is more prominent in western media outlets

<u>Experiment 2B:</u>

Number of all scores >0, 0, and 1 for eastern and western media can be displayed side-by-side to analyze whether any bias exists or not

<u>Result 2B:</u>

```
In [12]: bars1 = new_train[new_train['eastern/western']=='E'].groupby('Sign')['eastern/western'].count().tolist()
         bars2 = new_train[new_train['eastern/western']=='W'].groupby('Sign')['eastern/western'].count().tolist()
         barWidth = 1
         r = ['<0','=0','>0']
         plt.bar(r, bars1, color='red', edgecolor='white', width=barWidth, label = "Eastern")
         plt.bar(r, bars2, bottom=bars1, color='green', edgecolor='white', width=barWidth, label = "Western")
         plt.legend(loc = 0)

Out[12]: <matplotlib.legend.Legend at 0x25876847940>
```

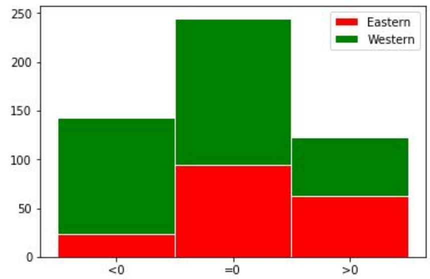


Figure 23: Result obtained from Experiment2B (./csv_analysis.ipynb)

There is a clear bias that can be seen from the graph. The numbers in category "<0" is disproportionately high for western media. Hence Claim 2B is **<u>correct.</u>**

<u>Claim 2C:</u> Sri Lanka's Hambantota Port deal to China is most talked about while portraying BRI in a negative light or example for "debt-trap" diplomacy. Another example used is Myanmar's renegotiation of loans to China and scaling down of a project. Mention of positives of BRI barely mention successful projects in a country, hence Sri Lanka and Myanmar should be often cited in news articles.

<u>Experiment 2C</u>:

Counting the total number of occurrences of a country in the corpus and counting the number of articles mentioning a particular country, relative to the other can possibly help resolve the claim.

<u>Result 2C:</u>

It can be observed from the graphs obtained that Sri Lanka, Pakistan, India, and Myanmar are the most mentioned and hence most important for analyzing data about BRI. (China is removed while calculating these graphs)

In [53]:
```
import operator
dict_occur = dict(sorted(dict_occur.items(), key=operator.itemgetter(1),reverse=True))
first10pairs = {k: dict_occur[k] for k in list(dict_occur)[:10]}

plt.bar(first10pairs.keys(),first10pairs.values())
plt.xticks(rotation=45)
plt.savefig("image2.jpg")
```

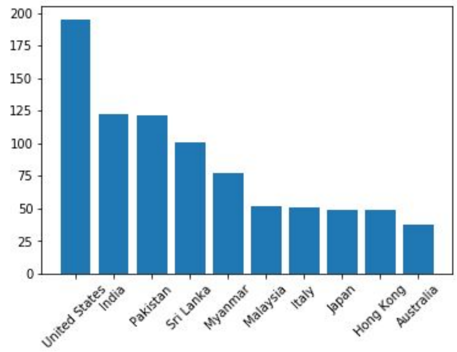


Figure 24: Occurance of country per article (./csv_analysis.ipynb)

In [78]:
```
first10pairs = {k: sorted_d[k] for k in list(sorted_d)[:20]}

plt.bar(first10pairs.keys(),first10pairs.values())
plt.xticks(rotation=45)
plt.savefig("image.jpg")
```

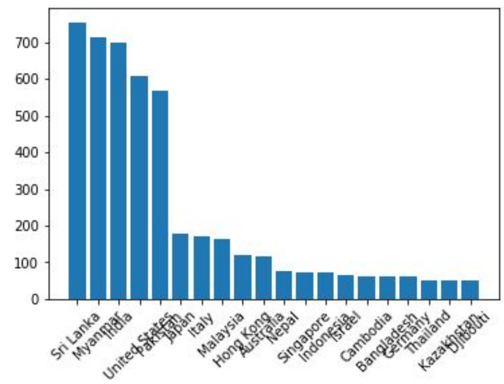


Figure 25: Occurance of country in corpus (./csv_analysis.ipynb)

Chapter 6 **Analysis of Result and Case Studies**

As observed from Figure 24 and 25, Sri Lanka and India is prominent while examining Belt and Road Initiative. Using data and analysis by Lowy Institute, AidData, Brookings Institute, Global Development Policy Center, Fitch Solutions, RWR Advisory, and Harvard Kennedy institution we would like to explore if our data is consistent with the theory and would like to draw a clearer portrait about ground reality of BRI in these countries. We consider 3 case studies- (a) Chinese loans in Africa, (b) India's rejection of BRI, and (c) Case of Hambanthota Port in Sri Lanka.

6.1 **Chinese loans in Africa**

One of the claims made is that China only invests in countries with rich resources and often create "white-elephant" projects or cause debt-asset swap.

Senior-level engagement in Africa by the Trump administration, which is critical to advancing commercial interests, has been virtually nonexistent [68] meanwhile China became one of the world's largest providers for short-term agricultural training courses [69]. In February 2020, Trump issued a travel ban on Nigeria (Africa's most populous country and having the largest market), Sudan, Tanzania, Libya, Somalia and Eritrea; which is over a quarter of the population of Africa.

Driven partially by economic transition at home, the BRI serves a range of domestic economic goals and geostrategic aims. BRI projects contribute to the internationalisation of Chinese infrastructure firms and directly benefit Chinese goods and exports—useful in offshoring excess capacity at home. For East and North Africa, this includes new maritime and transport infrastructure projects, including key segments of cross-border railway networks, special economic zones (SEZs), and industrial estates [70].

BRI signifies a shift in China's economic engagement with Africa, away from the resource trade that characterised the boom of the 2000s, towards a greater emphasis on infrastructure, industrial cooperation, and connectivity [70,72].

Djibouti: Regarded as BRI hub in Africa, it is also the location of the first overseas Chinese naval facility. China has financed multiple economic infrastructure projects totalling US$1.8 billion, including a multipurpose port and free trade zone complex at Doraleh [70]. China Telecoms has established a new data center in Djibouti that will connect it to other regional hubs in Asia, Europe, and to China, and potentially facilitate the development of submarine fibre cable networks in East Africa.

Debt repayment is tied to the sustainability—and profitability—of projects. This requires firstly, capacity on the part of African governments to adequately evaluate the returns to selected infrastructure projects, and that they are economically viable. Secondly, government agencies need domestic capacity building to be able to operate them in the long-term. This will entail continued involvement from Chinese enterprises in management and local training, but also pressure from host governments on foreign actors to facilitate technology and skills transfer towards African ownership.

In recent years the value of Chinese as well as US trade with Africa has declined due to fall in oil prices. Oil is still the top export product to both the countries (Angola is China's third largest oil partner). Recent studies [72,74] show that China engagement in Africa emphasizes infrastructure needs. The construction sector is a top destination for Chinese FDI stock in Africa, while transportation (roads, railroads, airports, and harbors) is a top destination for China Eximbank loans. Machinery, often used in infrastructure construction, is a top export to Africa for both China and the United States.

According to McKinsey [73], Chinese firms contribute to African economies—

- 89% of employees are African
- 64% of firms provide training
- 44% of managers are African

Another bias portrayed in media is that Chinese government engages more with less democratic government, which was proven to be false i.e. the relationships between aggregate governance quality and western and Chinese firms' respective levels of commercial engagement in Africa do not differ significantly [74].

The demand for debt in emerging from Africancountries, and external parties, China included, are not 'pushing' debt on Africa [75]. This is largely owing to the emergence of China as a major financier of African infrastructure, resulting in a narrative that China is using debt to gain geopolitical leverage by trapping poor countries in unsustainable loans. The dent-trap narrative also highly undermines African governments decision making powers.

Through research collected from various think-tanks and research organization, it is hence shown that narrative of debt-trap diplomacy and resource-hungry malicious intent in BRI is false.

6.2 India's rejection of BRI

India's rejection of BRI is skewed towards political reasons rather than financial reasons. As the BRI progresses, the Indian focus is more on pursuing its own connectivity plans (individually or with other partners). India's stance, from the earlier geopolitical and developmental aspects of the initiative, the focus is now shifting more towards a political economy analysis of participating countries. 2015). Four corridors, namely, the new Eurasia Land Bridge, China–Central Asia–West Asia Economic Corridor, the China– Pakistan Economic Corridor (CPEC) and the Bangladesh–China–India–Myanmar Economic Cooperation (BCIM), directly affects India's economic and strategic linkages with these regions.

Because of the overwhelming emphasis on the CPEC in Indian discussions (by Ministry of External Affairs), the perceptions were mainly shaped by geopolitical dimensions of the BRI rather than broader developmental aspects. Chinese are relaxed about the rise of India' but 'the Indians are much more nervous about the rise of China [76]. Geopolitical tensions between India and China are their shared but disputed border resulting in "Status quo ante bellum", a military stand-off that lasted more than 2 months. Officially, The Indian government has neither fully rejected the initiative or endorsed it in a clear manner [77] but takes a contradictory opposition of CPEC openly. CPEC would be passing through Kashmir, a region claimed by both India and Pakistan in its entirety, hence flagship project of OBOR reflects lack of appreciation of India’s concerns on the issue of sovereignty and territorial integrity. Despite not endorsing the BRI, New Delhi has participated in the Asian Infrastructure Investment Bank (AIIB) from the beginning. After China, India is now the second largest shareholder in the bank and also the largest recipient of concessional finance from the bank. Indian perceptions are also evolving which is stringtied under geopolitical dilemma.

6.2 Case of Hambantota Port in Sri Lanka

The sale of Hambantota Port took place on 29 July 2017. By creating graphs we can see that sentiment of articles have become more negative (i.e. towards “debt-trap” diplomacy). From year by year analysis of changing sentiment, we can see that western and eastern media is becoming more polarized towards the topic. Although it is hard to pin down as direct effect due to sale of port or news about various other projects as well.

```
In [22]: new_train["Year"] = year
sample_train = new_train[new_train['Year'] != '2009']
sample_train = sample_train[sample_train['Year'] != '2010']
sample_train = sample_train[sample_train['Year'] != '2014']
sample_train = sample_train[sample_train['Year'] != '0000']
print(len(sample_train))
sample_train.groupby('Year')['score'].count().plot.bar()
plt.show()
sample_train.groupby('Year')['score'].mean().plot.bar()
plt.show()
```

487

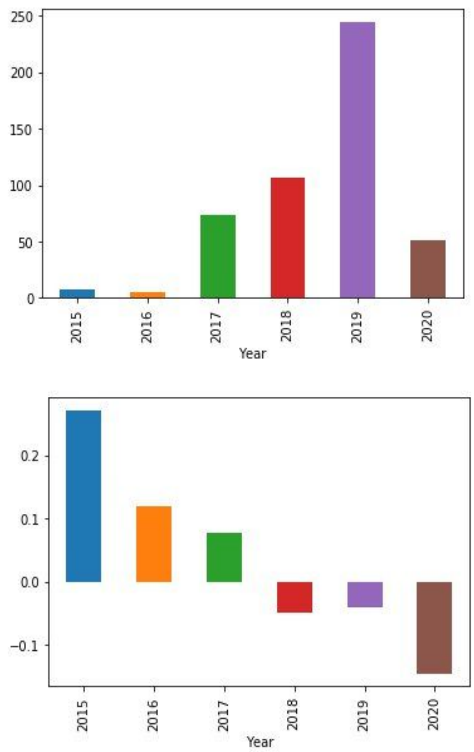


Figure 26: (A) Number of articles published/year (B)Average sentiment/year

```
In [71]: bars1 = sample_train[sample_train['eastern/western']=='E'].groupby('Year')['score'].mean().tolist()
         bars2 = sample_train[sample_train['eastern/western']=='W'].groupby('Year')['score'].mean().tolist()
         barWidth = 0.25
         bars1.insert(1,0)
         r1 = np.arange(len(bars1))
         r2 = [x + barWidth for x in r1]

         r = ['2015', '2016', '2017', '2018', '2019', '2020']
         plt.bar(r1, bars1, color='red', width=barWidth, edgecolor='white', label='Eastern')
         plt.bar(r2, bars2, color='green', width=barWidth, edgecolor='white', label='Western')

         plt.xticks([t + barWidth for t in range(len(bars1))], r)

         plt.legend(loc = 0)

         ['2017', '2015', '2020', '2018', '2019', '2016']

Out[71]: <matplotlib.legend.Legend at 0x258001b9710>
```

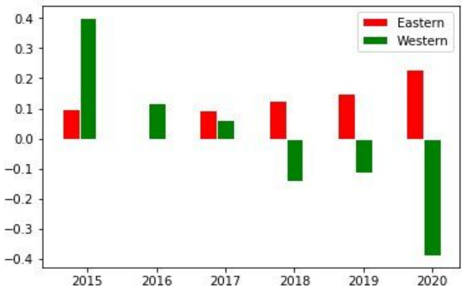


Figure 27: year by year analysis of average news sentiment

The economic reality is that Sri Lanka leased out Hambantota port to China largely due to a persistent balance of payment (BOP) crisis resulting from the reduction of trade over the years even while external debt servicing costs have been soaring. Sri Lank faced a severe shortage of foreign reserves in light of the upcoming debt servicing payments, due to the maturity of international sovereign bonds. Therefore, the country had to look for various avenues to obtain foreign currency inflows. Leasing out Hambantota port was one of the ways to increase the country's foreign reserves.

Although Hambantota port was leased to CM Port, the loans obtained to construct Hambantota port were not written off and the government is still committed to loan repayments as per the original agreements. The money obtained through leasing Hambantota port was used to strengthen Sri Lanka's dollar reserves in 2017-18, particularly in light of the huge external debt servicing due to the maturity of international sovereign bonds in early 2019.

By the end of 2017, only little over 10 percent of Sri Lanka's foreign debt was owed to China and most of that was in the form of concessionary loans. In addition to that, media reports have indicated that the government is planning to lease Mattala Rajapaksa International Airport (MRIA), one of the emptiest airports in the world, also located in Hambantota, to India.

On 25 June 2018, The New York Times published an article titled "**How China Got Sri Lanka to Cough Up a Port"** [84]. The article used an out-of-context quote from the President of Sri Lanka to The Hindu [83]. The interview had one question the president-

Q: You have said publicly you will renegotiate the Hambantota port agreement with China, which India was concerned about. Along with that is the future of Mattala airport, which India has shown an interest in. Now that you are in power, what will you do?

A: I believe that the Sri Lankan Government must have control of all strategically important projects like Hambantota. After all, these are not like a hotel or a terminal, but to give control of a port or an airport or our harbours is different.

With our control, they can do anything, but these 99-year lease agreements (that the previous government signed) will have an impact on our future. The next generation will curse our generation for giving away precious assets otherwise.

That is why our party protested these decisions.

This was quoted in the New York Times article highlighting debt-trap diplomacy. But the next immediate question in the interview was as follows-

Q: But the reason the lease had to be given was because of the debts incurred by the government of President Mahinda Rajapaksa...

A: No, that is wrong. It is also wrong to say there was a debt trap. In fact, during our time the ports authority paid back the first instalment (to Chinese banks). The Sirisena government, on the other hand, got more money as loans and just spent it. If they were worried about the debts piling up why didn't they first service the debt, rather than give away sovereignty?

In a rebuttal piece by Daily News Sri Lanka against The New York Times article was penned by Director of China Belt and Road Desk, Baker Tilly MH Advisory Sdn Bhd, Malaysia giving the following details (which was later on confirmed) -

- The first loan of $361M was offered by Ex-Im bank of China to fund 85% for Phase 1 of port after negotiations. It was a 15 year commercial loan @6.3% with short backed by collateral, 4 year moratorium. Quotas for preferential loan by Ex-Im bank had been used for Norochcholai power plant and other projects. Two options were offered to Sri Lanka for the interest rate (a) fixed 6.3% or (b) floating rate pegged to London Interstate offered rate which was 5pt at that time and rising. Sri Lanka chose the former.

- The feasibility study cited by NYT was done by Canadian company (rejected by ministerial task force) and another was done in 2006 by Danish company Ramboll giving positive results. **Both studies were done before the announcement of BRI in 2013.**

- **Hambantota Port suffered losses after opening**

  Construction work for Phase 1 of Hambantota Port, undertaken jointly by China Harbour Engineering Company (CMPH) and Sinohydro Corporation, commenced in January 2008. The port became operational on November 18, 2010, five months ahead of schedule. However, Hambantota Port was unable to generate sufficient revenue to meet its loan obligations due to inadequate governance, lack of commercial and industrial activities, as well as its inability to attract passerby vessels to dock at the port. By the end of 2016, it suffered a total loss of $304 million.

- **Majority control of Hambantota Port goes to CMPH**

  Amid mounting pressure to meet the International Monetary Fund's bailout terms and loan repayment obligations, the Sri Lankan government struck a public-private partnership (PPP) deal with China in July 2017, giving majority control of Hambantota Port to CMPH, which is listed on the Hong Kong Stock Exchange (HKSE).

  According to the filing made by CMPH to the HKSE, the terms of the concession agreement related to Hambantota Port were as follows:

  CMPH would make an investment of $1.12 billion in Sri Lanka, out of which $974 million would be used for the acquisition of 85 percent shares in the Hambantota International Port Group (HIPG), a company which was granted a 99-year term by the Sri Lankan government to develop, manage and operate Hambantota Port valued at $1.4 billion. HIPG would acquire 58 percent of Hambantota International Port Services (HIPS), which had been given the exclusive rights to develop, manage and operate the Common User Facility of Hambantota Port. The Sri Lanka Port Authority (SLPA) would hold 15 percent and 42 percent equity interest in HIPG and HIPS, respectively. The remaining $146 million would be deposited in CMPH's Sri Lanka bank account for the purpose of future development of the port and marine-related activities.

  Within 10 years from the effective date of the concession agreement, SLPA has the right to buy back 20 percent shares of HIPG on terms mutually agreed upon. After 70 years, SLPA could acquire CMPH's entire shareholdings in HIPG at a fair value to be determined by the valuers appointed by both parties. On expiry of 80 years, SLPA could buy up CMPH's shareholdings in HIPG for $1, leaving CMPH with 40 percent shareholdings in HIPH. After 99 years, CMPH would transfer all its shareholdings in HIPG and HIPS to the Sri Lanka government and SLPA at a token price of $1 upon termination of the agreement. The concession agreement went into effect on December 9, 2017. To increase industrial and commercial activities at the port, China further undertook to develop a 50 sq. km economic zone and build a liquefied natural gas plant and a tourist dockyard. China will also invest between $400 million to $600 million to develop phase 3 of Hambantota Port which is expected to be completed by 2021. The PPP thus is not a debt-equity swap but a fresh investment by CMPH amounting to $1.12 billion. The loan taken by SLPA for the construction of Hambantota Port was transferred to Sri Lanka's treasury. CMPH's investment in HIPG will be disbursed in three tranches of $292 million, $97 million and $584 million, with the balance of $146 million to be deposited in CMPH's Sri Lanka bank account for future use [82].

This example illustrates how western media sometimes deliver a bias. The case studies are not thorough. For policy making especially considering billions of dollars, complex analysis of policies, international relation, politics, environmental hazards etc. are scrutinized. It is unfair to pin it as "debt-trap" diplomacy without proper investigation.

Chapter 7 **Conclusion and Future Works**

The project was able to achieve the objectives successfully. The project was able to extract news article data and perform sentiment analysis and data analysis. The data along with policy research was analyzed to prove that indeed a media bias exists in the western media coverage of China's Belt and Road Initiative and further investigation should be conducted by journalists to provide a balanced and detailed analysis to his/her readers. The project was also able to prove through experiment about a pre-processing technique for sentiment analysis in a news article. This technique when clubbed with state of the art architecture such as [29] could give improved results.

For future directions, incorporating more media outlets in different countries can provide more accurate results. Additionally, TV media could also be added to the corpus to get a better understanding of public opinion and investigate its time series change.

The github for code used- https://github.com/tanu17/final-year-project

References

[1] Williams, A. N., and K. M. Woessner. "*Monosodium glutamate 'allergy': menace or myth?.*" Clinical & Experimental Allergy 39, no. 5 (2009): 640-646.
[2] Singh, Monica. "*Fact or Fiction? The MSG Controversy.*" (2005). Citable link: http://nrs.harvard.edu/urn-3:HUL.InstRepos:8846733
[3] Becker, Howard. "*The nature and consequences of black propaganda.*" American Sociological Review 14, no. 2 (1949): 221-235.
[4] Pacepa, Ion Mihai, and Ronald J. Rychlak. "*Disinformation: Former Spy Chief Reveals Secret Strategy for Undermining Freedom, Attacking Religion, and Promoting Terrorism.*" Wnd Books, 2013.
[5] Marwick, A. and Lewis, R., 2017. "*Media manipulation and disinformation online*". New York: Data & Society Research Institute.
[6] Lewis, David D., Yiming Yang, Tony G. Rose, and Fan Li. *"Rcv1: A new benchmark collection for text categorization research."* Journal of machine learning research 5, no. Apr (2004): 361-397.
[7] Myers, Daniel, and James W. McGuffee. "*Choosing scrapy.*" Journal of Computing Sciences in Colleges 31, no. 1 (2015): 83-89.
[8] Ed Kohlwey. "*yanytapi*". https://pypi.org/project/yanytapi/
[9] Ou-Yang, L., 2013. *Newspaper: Article scraping & Curation*. Python Library. Retrieved.
[10] Hamborg, F., Meuschke, N., Breitinger, C. and Gipp, B., 2017. "*News-please: a generic news crawler and extractor*". In 15th International Symposium of Information Science (ISI 2017) (pp. 218-223).
[11] Munson, Sean A., Stephanie Y. Lee, and Paul Resnick. *"Encouraging reading of diverse political viewpoints with a browser widget."* In Seventh International AAAI Conference on Weblogs and Social Media. 2013.
[12] Bernhardt, D., Krasa, S. and Polborn, M., 2008. *"Political polarization and the electoral effects of media bias."* Journal of Public Economics, 92(5-6), pp.1092-1104.
[13] Hall, Wendy, Ramine Tinati, and Will Jennings. *"From Brexit to Trump: Social media's role in democracy."* Computer 51, no. 1 (2018): 18-27.
[14] Goncalves, Jesse. *"Status Quo Bias in the Mainstream American Media Coverage of Senator Bernie Sanders."* SUURJ: Seattle University Undergraduate Research Journal 1, no. 1 (2017): 6.
[15] Mustafaraj, E., Lurie, E. and Devine, C., 2020, January. *"The case for voter-centered audits of search engines during political elections."* In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 559-569).
[16] Margaret Sullivan . "*The media keep falling in love — with anybody but Bernie Sanders*". The Washington Post, accessed 26 April 2020: www.washingtonpost.com/lifestyle/style/the-media-keep-falling-in-love--with-anybody-but-bernie-sanders/2020/02/12/0f55cc12-4d9c-11ea-bf44-f5043eb3918a_story.html
[17] Jon Allsop. *"Coverage of Bernie Sanders suffers from a lack of imagination"*. Columbia Journalism Review, accessed 23 April 2020: https://www.cjr.org/the_media_today/bernie_sanders_media_new_hampshire.php
[18] Robin Anderson. *"Factchecking NPR's Attempted Takedown of Bernie Sanders"*. FAIR (Fairness & Accuracy in Reporting), accessed on 7 March 2020: fair.org/home/factchecking-nprs-attempted-takedown-of-bernie-sanders
[19] Jake Johnson. *"Analysis of Primetime MSNBC Programs Finds Sanders Received 'Least' and 'Most Negative' Coverage of Top 2020 Democrats"* .Common Dreams.
[20] Sean A Munson, Daniel Xiaodan Zhou, and Paul Resnick. 2009. *"Sidelines: An Algorithm for Increasing Diversity in News and Opinion Aggregators"*. In ICWSM.

[21] Souneil Park, Seungwoo Kang, Sangyoung Chung, and Junehwa Song. 2009. *"NewsCube: delivering multiple aspects of news to mitigate media bias."* In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM,443–452.
[22] Souneil Park, Minsam Ko, Jungwoo Kim, H Choi, and Junehwa Song. 2011.*"NewsCube 2.0: An Exploratory Design of a Social News Website for Media Bias Mitigation"*. In Workshop on Social Recommender Systems.
[23] Brent H Baker, Tim Graham, and Steve Kaminsky. 1994. *"How to identify, expose & correct liberal media bias".* Media Research Center Alexandria, VA
[24] Matthew Gentzkow and Jesse Shapiro. 2005. *"Media bias and reputation."* Technical Report. National Bureau of Economic Research
[25] Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Vanni Zavarella, Erik Van Der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. 2013. *"Sentiment analysis in the news"*. arXiv preprint arXiv:1309.6202 (2013)
[26] S Shyam Sundar. 1999. *"Exploring receivers' criteria for perception of print and online news."* Journalism & Mass Communication Quarterly 76, 2 (1999), 373–386.
[27] Hamborg, F., Meuschke, N. and Gipp, B., 2017, June. *"Matrix-based news aggregation: exploring different news perspectives."* In 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL) (pp. 1-10). IEEE.
[28] Kim, Y., 2014. *"Convolutional neural networks for sentence classification."* arXiv preprint arXiv:1408.5882.
[29] Nguyen, Dinh, Khuong Vo, Dang Pham, Mao Nguyen, and Tho Quan. *"A deep architecture for sentiment analysis of news articles."* In International conference on computer science, applied mathematics and applications, pp. 129-140. Springer, Cham, 2017.
[30] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. *"Bert: Pre-training of deep bidirectional transformers for language understanding."* arXiv preprint arXiv:1810.04805 (2018).
[31] Akbik, Alan, Duncan Blythe, and Roland Vollgraf. *"Contextual string embeddings for sequence labeling."* In Proceedings of the 27th International Conference on Computational Linguistics, pp. 1638-1649. 2018.
[32] Medsker, L., Jain, L. "Recurrent Neural Networks: Design and Applications". CRC Press, Inc., Boca Raton (2001)
[33] Dor, Daniel. *"On newspaper headlines as relevance optimizers."* Journal of Pragmatics 35, no. 5 (2003): 695-721.
[34] Kirange, D. K., and Ratnadeep R. Deshmukh. *"Sentiment Analysis of News Headlines for Stock Price Prediction."* Composoft, An International Journal of Advanced Computer Technology 5, no. 3 (2016): 2080-2084.
[35] Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani. *"Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining."* In Lrec, vol. 10, no. 2010, pp. 2200-2204. 2010.
[36] Agarwal, Apoorv, et al. *"Opinion mining of news headlines using SentiWordNet."* 2016 Symposium on Colossal Data Analysis and Networking (CDAN). IEEE, 2016.
[37] Porter, M.F., 1980. *"An algorithm for suffix stripping."* Program, 14(3), pp.130-137.
[38] McCombs, Maxwell. *"New frontiers in agenda setting: Agendas of attributes and frames."* Mass communication review 24, no. 1/2 (1997): 32-52.
[39] McCombs, Maxwell E., and Donald L. Shaw. *"The evolution of agenda-setting research: Twenty-five years in the marketplace of ideas."* Journal of communication 43, no. 2 (1993): 58-67.
[40] Scheufele, Dietram A. *"Agenda-setting, priming, and framing revisited: Another look at cognitive effects of political communication."* Mass communication & society 3, no. 2-3 (2000): 297-316.
[41] Hester, Joe Bob, and Rhonda Gibson. *"The economy and second-level agenda setting: A time-series analysis of economic news and public opinion about the economy."* Journalism & Mass Communication Quarterly 80, no. 1 (2003): 73-90.
[42] Belt and Road Portal. Accessed on 1 May 2020: https://eng.yidaiyilu.gov.cn/home/rolling/85859.htm
[43] Y. LeCun, Y. Bengio, and G. Hinton, *"Deep learning,"* Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[44] M. A. Nielsen, *"Neural networks and deep learning,"* 2018.
[45] Fogel-Dror, Yair, Shaul R. Shenhav, Tamir Sheafer, and Wouter Van Atteveldt. *"Role-based Association of Verbs, Actions, and Sentiments with Entities in Political Discourse."* Communication Methods and Measures 13, no. 2 (2019): 69-82.

[46] "*Meeting Asia's Infrastructure Needs*". Asia Development Bank. Accessed on 3 May 2020: https://www.adb.org/sites/default/files/publication/227496/special-report-infrastructure.pdf
[47] Balahur, Alexandra, and Ralf Steinberger. *"Rethinking Sentiment Analysis in the News: from Theory to Practice and back."* Proceeding of WOMSA 9 (2009).
[48] Balahur Alexandra, Ralf Steinberger, Erik van der Goot, Bruno Pouliquen, Mijail Kabadjov (2009). *"Opinion Mining on Newspaper Quotations"*. Proceedings of the workshop 'Intelligent Analysis and Processing of Web News Content' (IAPWNC), held at the 2009 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology. Milano, Italy, 15.09.2009.
[49] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "*Bert: Pre-training of deep bidirectional transformers for language understanding."* arXiv preprint arXiv:1810.04805 (2018).
[50] International Monetary Fund. *"Internal Reserves and Foreign Currency Liquidity*." Accessed March 6, 2020.
[51] Van der Leer, Yeroen, and Joshua Yau. *"China's new silk route: The long and winding road."* PwC's Growth Markets Centre, February.(Accessed 7 September 2017) .
[52] Refinitiv. *"Belt and Road: An initiative in numbers"*. Accessed on 2 May 2020: https://www.refinitiv.com/content/dam/marketing/en_us/documents/reports/refinitiv-zawya-belt-and-road-initiative-report-2019.pdf
[53] Cheng, L.K., 2016. *Three questions on China's "Belt and Road Initiative"*. China Economic Review, 40, pp.309-313.
[54] Organisation for Economic Co-operation and Development (OECD) Report. *"China's Belt and Road Initiative in the global trade, investment and finance landscape"*. Accessed on 1 May 2020: www.oecd.org/finance/Chinas-Belt-and-Road-Initiative-in-the-global-trade-investment-and-finance-landscape.pdf
[55] Flint, Colin, and Cuiping Zhu. *"The geopolitics of connectivity, cooperation, and hegemonic competition: The Belt and Road Initiative."* Geoforum 99 (2019): 95-101.
[56] Liu, Weidong, and Michael Dunford. *"Inclusive globalization: Unpacking China's belt and road initiative."* Area Development and Policy 1, no. 3 (2016): 323-340.
[57] Yu Jie. "*China and Southeast Asia: Many Belts and Roads to Turn*".LSE IDEAS-CARI Special Report. October 2018.
[58] Bluhm, Richard, Axel Dreher, Andreas Fuchs, Bradley Parks, Austin Strange, and Michael J. Tierney. *"Connective financing: Chinese infrastructure projects and the diffusion of economic activity in developing countries."* (2018).
[59] Joanne Lu. "*China's Building Spree In Poor Nations: Does It Really Help The Local Economy?*". NPR.
[60] Hurley, John, Scott Morris, and Gailyn Portelance. *"Examining the debt implications of the Belt and Road Initiative from a policy perspective."* Journal of Infrastructure, Policy and Development 3, no. 1 (2019): 139-175.
[61] SUN, Degang, and Yahia ZOUBIR. *""Development First": China's Investment in Seaport Constructions and Operations along the Maritime Silk Road.*" Asian Journal of Middle Eastern and Islamic Studies 11, no. 3 (2017): 35-47.
[62] Cai, Peter. *"Understanding China's belt and road initiative."* (2017).
[63] Chitty, N, Ahlawat, D, Li, M & Gopal, D 2018, *'The Chinese Belt and Road Initiative and the Indian Ocean region: sentiment towards economic prosperity and security implications',* The Indian Journal of Politics, vol. 52, no. 1-2, 1, pp. 1-20.
[64] Brautigam, D., 2020. "*A critical look at Chinese 'debt-trap diplomacy': the rise of a meme"*. Area Development and Policy, 5(1), pp.1-14.
[65] Johnson, K. (2018, February 2). W*hy is China buying up Europe's ports?* Retrieved from https://foreignpolicy.com/2018/02/02/why-is-china-buying-up-europes-ports/
[66] Lee, J. (2018, November 29). *Beijing's Trojan Ports.* The American Interest. Retrieved from https://www. hudson.org/research/14717-china-s-trojan-ports
[67] Brautigam, D. (2019). *Misdiagnosing the Chinese infrastructure push*. The American Interest. Retrieved from https://www.the-american-interest.com/2019/04/04/misdiagnosing-the-chinese-infrastructure-push/
[68] Oxford Analytica. "*US influence in Africa will wane under Trump.*" Emerald Expert Briefings oxan-db.

[69] Tugendhat, Henry, and Dawit Alemu. *"Chinese agricultural training courses for African officials: Between power and partnerships."* World Development 81 (2016): 71-81.
[70] Chen, Yunnan. *"Silk road to the Sahel: African ambitions in China's belt and road initiative."* Policy Brief. China Africa Research Initiative 4 (2018).
[71] *"China Telecom Global Selects the Djibouti Data Center (DDC) as a Strategic Hub for Pan-African Expansion."* Marketwire: China Telecom Global, December 5, 2016.
http://www.marketwired.com/press-release/china-telecom-global-selects-djibouti-datacenter-ddc-as-strategic-hub-pan-african-expansion-2180862.htm.
[72] Janet Eom, Jyhjong Hwang, Lucas Atkins, Yunnan Chen, and Siqi Zhou. *"The United States and China in Africa: What does the data say?".* CAIR
[73] Yuan Sun, Irene, Kartik Jayaram, and Omid Kassiri. *"Dance of the lions and dragons: how are Africa and China engaging, and how will the partnership evolve?."* (2017).
[74] Landry, David G. *"Comparing the Determinants of Western and Chinese Commercial Ties with Africa."* (2019).
[75] Were, Anzetse. *"Debt trap? Chinese loans and Africa's development options."* (2018).
[76] Grant, Charles. *"India's response to China's rise."* (2009).
[77] Sachdeva, Gulshan. *"Indian perceptions of the Chinese Belt and Road initiative."* International Studies 55, no. 4 (2018): 285-296.
[78] Carrai, Maria Adele. *"China's Malleable Sovereignty along the Belt and Road Initiative: The Case of the 99-Year Chinese Lease of Hambantota Port."* NYUJ Int'l L. & Pol. 51 (2018): 1061.
[79] Roy-Chaudhury, Shantanu. *"China, the Belt and Road Initiative, and the Hambantota Port Project."* St Antony's International Review 15, no. 1 (2019): 153-164.
[80] Chitty, Naren, Dalbir Ahlawat, Mei Li, and D. Gopal. *"The Chinese Belt and Road Initiative and the Indian Ocean region: sentiment towards economic prosperity and security implications."* The Indian Journal of Politics 52, no. 1-2 (2018): 1-20.
[81] *"The 19th amendment is a failure" – President Rajapaksa.* The Daily News. Accessed on 2 May 2020:
https://www.dailynews.lk/2019/12/02/features/204423/%E2%80%9C-19th-amendment-failure%E2%80%9D-%E2%80%93-president-rajapaksa
[82] *Reply to NYT article published on China Focus Did China get Sri Lanka to cough up a port?.* The Daily News. Accessed on 2 May 2020:
https://www.dailynews.lk/2018/08/06/features/158855/did-china-get-sri-lanka-cough-port
[83] *Will be frank with New Delhi to avoid misunderstandings: Gotabaya Rajapaksa.* The Hindu. Accessed on 2 May 2020:
https://www.thehindu.com/news/international/need-more-coordination-between-delhi-colombo-says-gotabaya-rajapaksa/article30125809.ece
[84]*How China Got Sri Lanka to Cough Up a Port.* The New York Times. Accessed on 2 May 2020:
https://www.nytimes.com/2018/06/25/world/asia/china-sri-lanka-port.html