

Uncovering the impact of COVID-19 on shipping and logistics

Impact of
COVID-19 on
shipping and
logistics

Enna Hirata

Center for Mathematical and Data Sciences, Kobe University, Kobe, Japan, and

Takuma Matsuda

Takushoku University, Tokyo, Japan

305

Received 11 March 2021
Revised 26 May 2021
30 September 2021
8 October 2021
Accepted 8 October 2021

Abstract

Purpose – This research aims to uncover coronavirus disease 2019's (COVID-19's) impact on shipping and logistics using Internet articles as the source.

Design/methodology/approach – This research applies web mining to collect information on COVID-19's impact on shipping and logistics from Internet articles. The information extracted is then analyzed through machine learning algorithms for useful insights.

Findings – The research results indicate that the recovery of the global supply chain in China could potentially drive the global supply chain to return to normalcy. In addition, researchers and policymakers should prioritize two aspects: (1) Ease of cross-border trade and logistics. Digitization of the supply chain and applying breakthrough technologies like blockchain and IoT are needed more than ever before. (2) Supply chain resilience. The high dependency of the global supply chain on China sounds like an alarm of supply chain resilience. It calls for a framework to increase global supply chain resilience that enables quick recovery from disruptions in the long term.

Originality/value – Differing from other studies taking the natural language processing (NLP) approach, this research uses Internet articles as the data source. The findings reveal significant components of COVID-19's impact on shipping and logistics, highlighting crucial agendas for scholars to research.

Keywords Web mining, Corona virus, COVID-19, Text mining, Machine learning, Natural language processing, NLP, Shipping, Logistics

Paper type Research paper

1. Introduction

The coronavirus disease 2019 (COVID-19) pandemic has been affecting over 200 countries and territories across all regions. The pandemic is undoubtedly applying pressure on global manufacturer production capacities and logistics as an integral part of supply chains. Countries and companies have been taking various actions to protect the delivery of essential materials, including foods, medicines, masks and hazmat suits. While many countries have successfully managed the pandemic in phases, some logistics activities are gradually resuming to the prior pandemic scale.

Nonetheless, the crisis has led to rapid deterioration of business indicators, including GDP and productivity, and the impact on the world economy is predicted to be 2–3 times more severe than the world financial crisis in 2008–2009 (Harris, 2020). In the longer run, productivity is likely to be reduced by diminished R&D expenditure and diverted resource allocation of senior management to deal with the pandemic (Bloom *et al.*, 2020). In addition, Brinca *et al.* (2020) measure the impact of COVID-19 on the labor demand and suggest that the demand for working hours decreased by 16% due to the imposition of travel and trade restrictions and the shutdown of working places. The authors conclude two-thirds of the fall in the growth rate of hours worked in March and April 2020. It could be attributed to adverse labor supply shocks,



© Pacific Star Group Education Foundation. Licensed re-use rights only.

Declaration of Interests: none.

Funding: This research was funded by the Japan Society for the Promotion of Science (JSPS KAKENHI), grant numbers 20K22129, 21H01564.

Maritime Business Review
Vol. 7 No. 4, 2022
pp. 305-317
Emerald Publishing Limited
2397-3757
DOI 10.1108/MABR-03-2021-0018

and it suggests that correctly measuring demand and supply shocks is essential for the design and implementation of economic policy during the COVID-19 outbreak.

Logistics companies, which are involved in the transport, storage and delivery of goods, have been directly affected by the COVID-19 pandemic. COVID-19 changed the way to connect between institutions, companies and individuals physically. However, the impact of COVID-19 on shipping and logistics is difficult to measure, and some of them are yet to be observed.

This research aims to gain insights through semantic analysis of Internet articles that discuss the impact of COVID-19 on shipping and logistics. This research presents a method of applying natural language processing (NLP) to extract Internet articles. The extracted text documents are then trained by machine learning (ML) algorithms which perform automatic text classification. We collected articles from Internet resources in viewing that the literature in this area is yet scarcely available in scientific journals. The information extracted from online resources is then analyzed by the NLP method and ML algorithms for usual insights.

The rest of the paper is organized as follows: [Section 2](#) reviews the literature; [Section 3](#) discusses the methodology of NLP and ML algorithms; [Section 4](#) outlines the results of ML in visual style and discusses the research findings; [Section 5](#) conveys concluding remarks and prospects.

2. Literature review

2.1 COVID-19 impact

Since the outbreak of COVID-19 in late December 2019, governments and organizations in affected countries have been taking countermeasures relentlessly. The impacts of COVID-19 on supply chains are continuously measured, modeled and visualized. The literature exists on modeling the outbreak with individual and governmental actions, including holiday extension, city lockdown, hospitalization, and quarantine ([Lin et al., 2020](#)) and measuring pandemic and lockdown impact on mental health ([Rossi et al., 2020](#)), etc. Some scholars have pioneered in researching the impact of COVID-19 supply chains: [Choi \(2020\)](#) builds an analytical model and suggests that mobile service operation (MSO) is a win-win model for both the service provider and consumer. [Govindan et al. \(2020\)](#) apply the fuzzy inference system (FIS)-based model to help with the demand management in the health-care supply chain. [Ivanov \(2020\)](#) conducts a simulation study and suggests that the timing of the closing and opening of the facilities at different echelons might become a significant factor determining the epidemic outbreak impact on supply chain performance. While the impacts are multi-faceted, we mainly reviewed the literature in three streams as follows:

The first stream discusses the pandemic and supply chain resilience (SCR). Globalization provides a wealth of new opportunities for supply chain optimization and exposes supply chain networks to disruptions associated with increased complexity ([Golan et al., 2020](#)). The authors also state that impact of the COVID-19 pandemic on supply chains may be delayed with waves that can continue for months and even years; collaboration with nontraditional supply chain stakeholders plays a role in increasing SCR during the pandemic outbreak. [Kawasaki et al. \(2015\)](#) also mentioned that border impeded cargo flow and international trade. In particular, its border resistance increases in times of emergency, such as the COVID-19 outbreak. That was one of the reasons to reduce the trade volume of China temporarily.

The second stream discusses the pandemic and supply chain sustainability (SCS). [Karmaker et al. \(2021\)](#) investigate the drivers of SCS in the context of the COVID-19 pandemic in Bangladesh and suggest that financial support from the government and supply chain partners is required to tackle the immediate shock on SCS. The pandemic in the short term enables environmental sustainability gains, while long-term effects are uncertain ([Sarkis, 2020](#)).

The third stream discusses the pandemic and disruption management. Supply chain disruptions, such as the financial crisis in 2008 and the tsunami that hit Japan in 2011, have been known to cause significant challenges and affect an organization's performance ([Hendricks and](#)

Singhal, 2003). Mahajan and Tomar (2020) investigate the distance to production zones from our retail centers and find that long-distance food supply chains have been hit the hardest during the pandemic with welfare consequences for urban consumers and farmers.

2.2 Natural language processing (NLP) and its application in assessing the COVID-19 impact

NLP, also known as text mining, is a branch of artificial intelligence that helps computers understand, interpret and manipulate human language. NLP helps computers communicate with humans in their language and scales other language-related tasks. For example, “NLP makes it possible for computers to read text, hear speech, interpret it, measure sentiment and determine which parts are important. Today’s machines can analyze more language-based data than humans, without fatigue and in a consistent, unbiased way” (SAS, 2021). Automation through ML is critical to comprehensively analyze text and speech data because of the staggering amount of unstructured data generated every day, from medical records to social media.

From academia to industry, text mining has become a popular strategy for keeping up with the rapid growth of information. Automatic text mining methods can make extracting information from a large set of documents more efficient. However, since computer programs do not quickly process natural language, algorithms must be developed to transform texts into a structured representation.

From a data processing point of view, words or elements are “tokens” that provide context to language, clues to the meaning of words and those words’ relationships with other words. ML allows machines to use those tokens to identify the true meaning of what is said. ML is a subset of artificial intelligence (AI). The relationship of NLP versus other ML methodologies and the positioning of this research can be outlined as follows (see Figure 1):

Serrano *et al.* (2020) collect user comments on YouTube videos and apply NLP and other ML techniques to predict COVID-19 misinformation. Jelodar *et al.* (2020) take topic modeling and the recurrent neural network approach to analyze the online discussion forum Reddit to uncover various issues related to COVID-19. Hosseini *et al.* (2020) apply NLP to analyze the contents of tweets during the COVID-19 pandemic.

To the best of our knowledge, research extracting information from Internet articles to assess the COVID-19 impact on logistics has not been observed. This paper hence aims to complement the existing literature and offers insights on future research directions and focuses. The research extracts information from online resources, and applies the NLP method and ML algorithms to analyze the information extracted.

The research questions (RQs) are designed as follows:

- RQ1. What key topics are being discussed in terms of COVID-19’s impact on shipping and logistics?
- RQ2. What areas should researchers and policymakers give priority to?

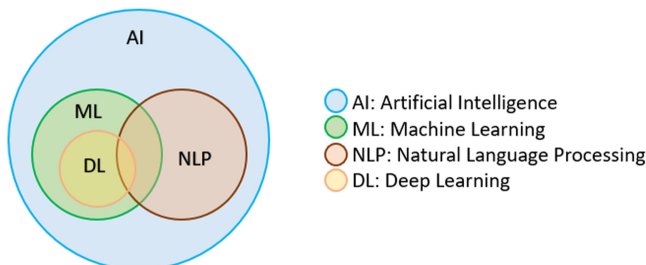


Figure 1.
Relationship between
NLP, ML and AI

3. Methodology

The proposed method is organized into four major modules: data processing, information interaction, ML and visualization. The data processing stage involves the techniques and processes which conduct tasks of text mining. In the information extraction stage, we use word2vec, a natural language processing method that converts words in a text into numerical vectors to understand their meanings. Then principal components analysis and *t*-distributed stochastic neighbor embedding (*t*-SNE) are used for clustering. Finally, the visualization phase describes the findings of the study. The workflow of the proposed system is presented in Figure 2.

3.1 Data preprocessing

The data are collected from articles on the Internet using the web-mining method. A wide variety of works in the literature focused on text mining for web contents mining (Singh and Singh, 2010). Web mining is the application of data-mining techniques to discover patterns from the World Wide Web. It retrieves structured and unstructured information from browser activities, server logs, website link structures, page contents and other sources. The benefit of web mining is that it allows users to quickly discover useful information or knowledge from the Internet website structure, which otherwise requires human work to collect manually.

Web mining includes web content mining, web structure mining and web user information mining. The goal of web structure mining is to generate structural information from websites and webpages. In this research, we focus on the structure of inner-document to get the web contents as text documents for analysis.

The contents retrieved from web mining are further analyzed through a series of NLP text analysis methods. Text analysis allows automatic extraction and classification of information from texts (e.g. Westergaard *et al.*, 2018), such as tweets, emails, support tickets, product reviews, survey responses and web contents. Popular techniques in text analysis include word frequency, collocation, co-occurrence, text classification, sentiment analysis, topic detection, language detection, clustering, keyword extraction and entity recognition. Sorting through data could be repetitive, time-consuming and expensive if done by humans. Instead, if done by machines, high volumes of text can automatically be analyzed, saving resources while providing broader insights.

The web and text mining in this study is performed using the following steps:

3.1.1 Corpus generation. The experiment of this study is carried out on a text corpus, a collection of Internet articles. The articles are retrieved by using the keywords “COVID corona virus impact on shipping and logistics.” Total 283 websites hit in the Google search on

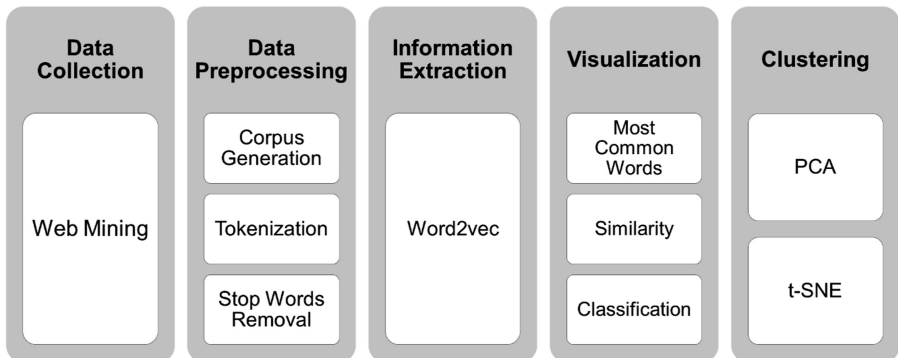


Figure 2.
Stages of the analysis

June 14, 2020. We focused on English articles in this study. Our categorization for industries and areas of websites are shown in Table 1. “Shipping, Port” indicates shipping companies and port operators and port authorities. “Logistics” means forwarders, truckers and so on. “ICT” includes software companies, logistics solution providers. “Shippers” are manufacturers and retailers. “Finance, Insurance” indicates financial institutions and insurance companies such as Protection and Indemnity (P&I) company. “Equipment” is a ship equipment maker. “Public and educational institutions” includes universities, research institutions, public sectors in countries and international institutions. “Media” means newspaper companies and Internet news providers, the latter tend to engage in shipping and logistics supply-chain matters. Many of the extracted websites are reports and articles describing the impact of COVID-19 on logistics.

The web mining is performed by using python programs (beautifulsoup and pdfminer) to convert web information to text format for further analysis.

3.1.2 Tokenization. Tokenization is a critical and the most basic step to proceed with NLP. In NLP, tokenization refers to splitting a phrase, sentence, paragraph or an entire text document into smaller units, such as individual words or terms. Each of these smaller units is called a token. These tokens are the key elements of the NLP.

Tokenization is the process to replace the meaningful sentence in to individual words with space as the delimiter, while remaining all the valuable information. As an example, let us consider the below sentence:

COVID-19 is proving to be among the greatest of logistics disruptors in modern times

Tokenizing the sentence, we will get:

["COVID-19", "is", "proving", "to", "be", "among", "the", "greatest", "of", "logistics", "disruptors", "in", "modern", "times"]

The python library of Natural Language Tool Kit (NLTK) is applied in this study to split the text data to tokens. Once sentences are tokenized, the next step is to clean the text by removing stop words to get ready for the model building part.

3.1.3 Stop words removal. The next stage of data preprocessing is stop words removal. Stop words are words which are filtered out before or after processing of natural language data. Stop words are the most common words in any natural language. For analyzing text data and building NLP models, those stop words might not add much value to the meaning of the document, so they are usually removed to improve ML accuracy.

Industry/area	Africa	Asia	Europe	Oceania	South America	North America	World or n.a	Total
Shipping, port		1	4	1		3		9
Logistics	1	16	25	1	1	30		74
ICT		5	10	1		8		24
Shippers		4	6			7		17
Finance, insurance		1	2			1		4
Equipment		1	1					2
Public and educational institutions	1	1	5			8	4	19
Media		13	33			49	9	105
Consulting/advisory	3	1	15	1		6	1	27
n.a							2	2
Total	6	43	101	4	1	112	16	283

Table 1.
Industries and areas of data source

Stop words usually refer to the most common words in a language; however, there is no single universal list of stop words. In English, it includes words like “is”, “was”, “were”, “the”, “a”, “for”, “of”, “in”, etc. In this study, the NLTK English stop word set is applied to remove common stop words in English. We also removed some extra stop words (e.g. “https”, “www”, “TOP”, “BACK”, etc.) that are particularly associated with the corpus. Removing stop words helps reduce the size of the corpus and identify the keywords in the corpus as well as frequency distribution of concept words in the overall context.

After removing stop words, the tokenized sentence in the above example contains:

[“COVID-19”, “proving”, “greatest”, “logistics”, “disruptors”, “modern”, “times”]

In addition, to unify the expression for best accuracy, we replaced “corona”, “corona-virus” and “covid” with “covid-19”.

After data preprocessing, we have enough tokenized clean text for the machine to work with, and to develop algorithms to differentiate and make associations between pieces of text to make predictions.

The input and output of a sample sentence in each process is summarized in [Figure 3](#).

3.2 Information extraction and machine learning

Information is extracted by applying a couple of ML algorithms. ML is a type of AI that allows systems to improve (learn) self-driving through experience. ML algorithms build a mathematical model based on sample data, known as “training data,” to make predictions or decisions without being explicitly programmed to perform the task. ML algorithms are used in various applications, such as email filtering and computer vision, where it is difficult or infeasible to develop a conventional algorithm for effectively performing the task. This study first applies Word2vec to vectorize the text information, followed by principal component analysis (PCA) and *t*-SNE algorithms for dimension reduction and visualization.

3.2.1 Word2vec. Word2vec is applied to extract information from the corpora. [Mikolov et al. \(2013\)](#) propose the word2vec model for computing continuous vector representations of words from massive data sets, and observe considerable improvement in the quality of these representations measured in a word similarity. Word2vec applies a two-layer neural network deep learning model that processes text by vectorizing words. To be more specific, it first constructs a dictionary of words from the training text data and then learns vector representation of those words; as such, it turns text into a numerical form that deep neural networks can understand.

Word2vec conducts semantic comparisons ([Mikolov et al., 2013](#)) ranging from country–capital (e.g. “Poland” is to “Warsaw” as “Japan” is to “Tokyo”) and male–female (e.g. “man”

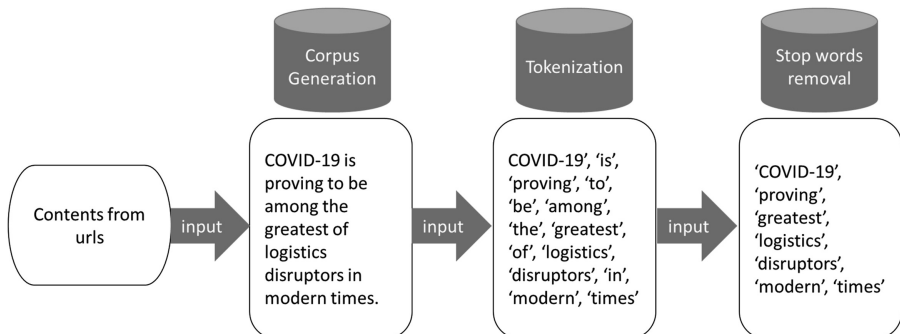


Figure 3.
A demonstration of data preprocessing

is to “son” as “woman” is to “daughter”). Word2vec can make highly accurate guesses about a word’s meaning based on past appearances given enough data, usage and contexts. Those guesses can be used to establish a word’s association with other words (e.g. “man” is to “king” what “woman” is to “queen”) or cluster documents and classify them by topic. Those clusters can form the basis of search, sentiment analysis, and recommendations in diverse fields such as scientific research, legal discovery, and customer relationship management. The training algorithm applied in this research is CBOW (Continuous Bag of Words). CBOW architecture tries to predict the current target word (i.e. the center word) based on the surrounding words.

The usefulness of Word2vec is that it detects similarities of words mathematically. Inputs are raw texts; Word2vec then converts words to vectors that are distributed numerical representations of word features such as the context of those words, respectively. The output of the Word2vec neural net is a vocabulary in which each item has a vector attached to it, which can be fed into a deep-learning net or queried to detect relationships between words.

3.2.2 Principal component analysis (PCA). PCA is prevalent in many academic areas such as psychology, sociology, civil engineering, etc. In recent years, PCA also has turned to be used as one of the ML algorithms. It is one of the multivariate analyses that uses an orthogonal transformation that converts a set of correlated variables to a set of uncorrelated variables, that is, to reduce the dimensions of the data. When there are many variables (e.g. features $n > 10$), it is advisable to do PCA. PCA is the most widely used tool in exploratory data analysis and ML for predictive models. PCA is also an unsupervised learning technique to examine the interrelations among a set of variables. It is also known as a general factor analysis where regression determines a line of best fit.

PCA is fundamentally a dimensionality reduction algorithm, but it can also be helpful as a tool for visualization, noise filtering, feature extraction, and engineering, and much more. As there are as many principal components as there are variables in the data, principal components are constructed so that the first principal component accounts for the largest possible variance in the data set.

Using PCA, we can capture the characteristics of a word with as few variables as possible by creating variables that combine the numerical values indicated by each vector element appropriately. To do this, we first look for an axis that maximizes the variance of the projected data so that the loss of information in the original data is as small as possible. It is called the first principal component. Then, among the axes orthogonal to the first principal component, find the axis that maximizes the variance of the projected data on the axis, and call this the second principal component. It will reduce the dimension of the data. The magnitude of the variance of the projected data is evaluated by the cumulative contribution rate (CCR) and is defined by the maximal amount of variance explained by the principal components representing the directions of the data. It also means the percentage of information accounted for by the selected principal component. In this way, we can measure the amount of information we lose due to dimensional shrinkage.

We use PCA to analyze vectors generated by word2vec. PCA reduces the data dimension avoiding information loss based on the CCR. Then we construct axes pc1 and pc2 based on the first and the second principal component, respectively. Finally we regenerate vectors for words.

3.2.3 *t*-distributed stochastic neighbor embedding (*t*-SNE). *t*-SNE is also an ML algorithm to reduce the data dimension. It is characterized by a reduction in the dimension so that the distances between data represented by low-dimensional vectors match the distances between data originally represented by high-dimensional vectors. [Van der Maaten and Hinton \(2008\)](#) proposed *t*-SNE as a tool to visualize high-dimensional data. Stochastic neighbor embedding ([Hinton and Roweis, 2002](#)), also known as SNE, converts the high-dimensional Euclidean distances between datapoints into conditional probabilities representing similarities. SNE constructs reasonably good visualizations, and sometimes it is challenging to optimize constrained by the cost function it utilizes. [Van der Maaten and Hinton \(2008\)](#) proposed *t*-SNE

by using a student's t distribution in the low dimensional space, which alleviates the problem that SNE suffers.

The efficiency of t -SNE is measured by the Kullback–Leibler (KL) divergences score. The KL score is calculated as the negative sum of probability of each event in P multiplied by the log of the probability of the event in Q over the probability of the event in P . A lower KL score indicates the higher performance of the t -SNE solution.

We also use t -SNE to analyze vectors generated by word2vec. It reduces the data dimension avoiding information loss based on the KL score. Then we construct axes ts1 and ts2. Finally we regenerate vectors for words.

4. Visualization of research findings

This section describes the findings from the ML in [Section 3](#).

4.1 Most common words

The most commonly used words in the research data are the output in [Figure 4](#). The word that appears most frequently in the corpus is “covid,” followed by “service,” “country,” “company,” and “China” in that order.

A broader view is computed by generating a word cloud ([Figure 5](#)). The bigger and bolder the word appears, the more often it is mentioned in the corpus.

4.2 Word similarity

Word similarity is measured by using cosine similarity, which is the cosine of the angle between two vectors generated by word2vec. No similarity is expressed as a 90-degree angle with a similarity of zero, while the total similarity of one is a 0-degree angle, complete overlap. As shown in the results of 4.1, the occurrence of the word “COVID” was the most prominent in both the frequency and the word cloud. Therefore, we would like to examine the similarity between “COVID” and other words. [Table 2](#) is a list of words associated with “COVID” using output from the Word2vec model, in order of proximity:

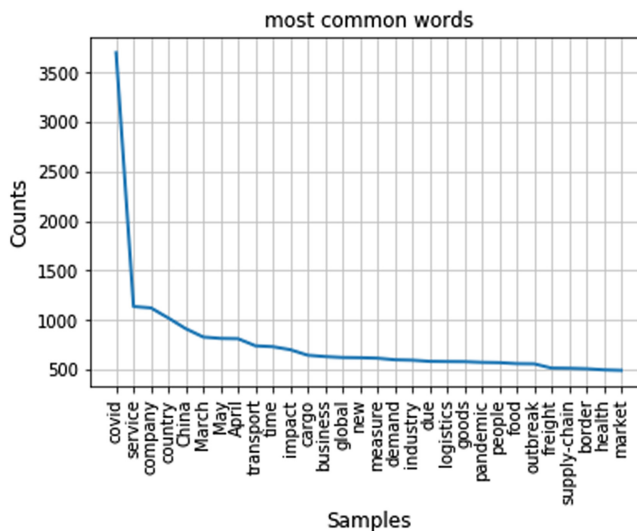


Figure 4.
Plot of most
common words



Figure 5.
Word cloud

Word	Similarity	Word	Similarity
Pandemic	0.951578	Business	0.642269
People	0.887602	Industry	0.632889
Food	0.882532	Demand	0.626850
Impact	0.842956	Measure	0.621253
Outbreak	0.836285	Country	0.555152
Company	0.740543	Global	0.551881
April	0.725649	May	0.539259
March	0.693238	Supply-chain	0.525226
New	0.666996	China	0.437704
Time	0.655033	Logistics	0.408916

Table 2.
Word similarity
(top 20)

The implications are three-fold:

- (1) “Pandemic” proves to be the most similar word to “COVID-19”. “People” and “food” are also considered most critical in relating to the COVID-19 pandemic.
- (2) “March,” “April,” and “May” are listed in the top 20 of most similar words, while “April” is more similar than “March” and “May.” This implies that the concerns on COVID-19’s impact on logistics peaked in April. Relevant parties may be better prepared in May in tasking/remedying COVID-19 disruption.
- (3) “China” is also in the list of top 20 most similar words, while “mask” is not even showing in the top 200 most common words.

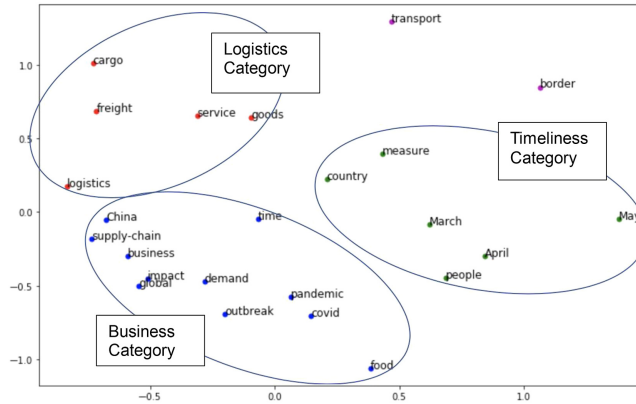
4.3 Word classification by PCA and t-SNE

Text classification is the task of assigning a set of categories to free text. Text classifiers can be used to organize, structure and categorize words. For example, chat conversations can be organized by language, and brand mentions can be organized by sentiment, and so on.

By reducing the dimensions of word vectors using PCA, the most common words are classified into three components. And the CCR was 0.84, indicating that the PCA model is efficient.

Figure 6 indicates the classification result; horizontal axis is the first principal component, vertical axis is the second principal component. The results from the PCA model can be summarized into three categories and a few outliers:

Figure 6.
Word classification
by PCA



- (1) Category related to logistics. The words “cargo,” “freight,” “logistics,” “service” and “goods” are associated with each other closely.
- (2) Category related to timeliness. The words “country,” “measure,” “people,” “March” and “April” are associated with each other closely.
- (3) Category related to business. “outbreak,” “pandemic,” “COVID,” “food,” “China,” “supply-chain,” “demand” and “global” are associated with each other closely.
- (4) Category of outliers. The words “transport” and “border” are outliers, which may explain that these are popular topics; however, they are not regarded as topics that are mainly related to COVID-19.

The result of *t*-SNE with the lowest KL score is plotted in [Figure 7](#). The words can be classified into three groups. Group (a) is related to the international transport of goods. Group (b) is related to the pandemic and its impacts on people’s work and life. Group (c) impacts on the world supply chain. The word “US” is classified in group (b), which indicates that the new norm of work and life was extensively discussed in the USA during the period. The word “China” is classified in group (c), which indicates that China plays a role in the global supply chain in various perspectives, including production, business, demand, shipping, logistics, etc.

As can be seen from [Figures 6 and 7](#), the different dimension contraction methods result in different groupings. It may be because PCA mainly focuses on capturing the linear structure of the data, while *t*-SNE captures both linear and nonlinear relationships ([Yang et al., 2017](#)). Although it is not easy to interpret the coordinates for PCA, *t*-SNE is regarded to be better than PCA in terms of visualization (e.g. [Yang et al., 2017](#)). PCA results clearly indicate that “Business”, “Logistics” and “Timeliness” are most important categories during the pandemic, which reflects perceptions in real life; therefore, we decided to report it for the readers’ perusal. Discussion in the concluding remarks is mainly based on *t*-SNE.

5. Conclusion remarks and future work

In the research, we apply web mining and NLP methods to perform large-scale analyses on Internet users’ opinions from different perspectives. Through automatic manipulation of Internet articles addressing COVID-19 using ML models, this research took an empirical approach to gain insights supported by statistical methods of ML algorithms.

The research findings can be summarized in three points. First, the impact of COVID-19 on shipping and logistics is that information on the sustainability of the service is one of the most important ones. In the similarity analysis, “measure” was high. It seems to result from the fact

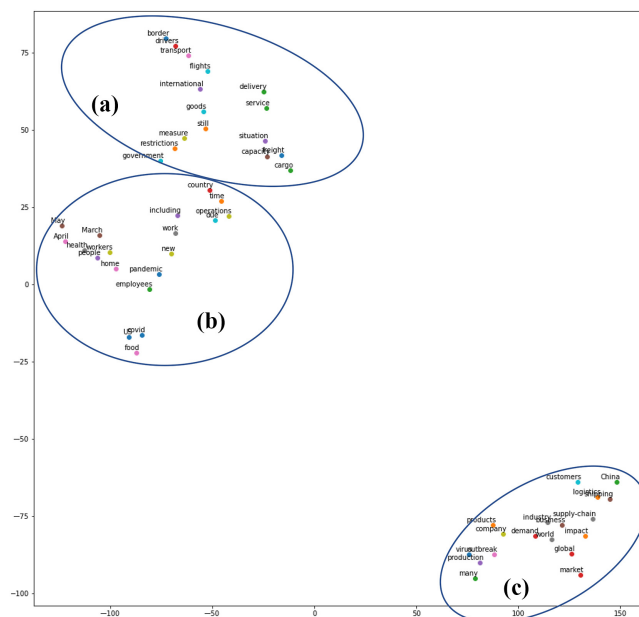


Figure 7.
Word classification by
t-SNE

that logistics companies focus on how transportation will be affected in the face of the spread of the infection and how to secure alternative means. Group (c) in Figure 7 also showed a high level of interest in “service,” “border,” “restrictions,” and “delivery”. In Figure 7, group (c), “service,” “border,” “restrictions,” “delivery,” and other topics related to the sustainability of services dominate the list. It is related to the fact that most of the extracted information sources on the Internet are related to logistics companies and logistics services, which are on the supply side, and are considered to reflect their needs.

Second, there is an emphasis on information related to consumers’ lives. In particular, “food” tends to be the focus of attention. It indicates that consumers are concerned about the effects of COVID-19 on their lives, such as food and health. Understandably, information providers in web news or logistics companies tend to reflect this concern. In the similarity analysis, “Food” was the only item in the top 20 items transported, reflecting its high level of attention. Since consumers tend to value information on the availability of daily necessities, supplier-side companies and trade magazines, which were the primary sources of this survey, cannot ignore this information. In Group (b) of Figure 7, there are many terms such as “health,” “home,” and “work” that are rooted in consumers’ daily lives. It suggests that it is important for the logistics business to know the situation and changes in consumers’ lives and food. The fact that words indicating the timing are in this group is because consumers are likely to emphasize transportation delays.

Third, topics related to “service,” “country” and “China” dominate the list. It is not unrelated that the extracted Internet information sources were mainly from Asia, Europe and North America. In Asia, information related to China was important as an export base for containerized cargo, while in Europe and North America, information related to China was important as an import source. In the similarity analysis, “April” was higher than the other months because the peak production stoppage and export decrease in China was in April 2020. In Figure 7, group (c), there are many terms around China, such as “production,” “market” and “supply-chain,” which are often used by producers and transporters. It may be a

result of the fact that the impact of COVID-19 has focused attention on the impact on supply chains related to China.

In addition, in the context of the findings of this study, it makes sense for researchers and policymakers to focus on the following two aspects:

- (1) Ease of cross-border trade and logistics. The word “border” has been a highly discussed topic; cross-border policy seems to be one of the critical logistics enablers during the pandemic. For this purpose, providing information on the sustainability of services from time to time would help facilitate the services. In addition, for smooth international logistics, the digitization of supply chains and the application of innovative technologies such as blockchain and IoT are required more than ever. It seems helpful to provide information on how the sustainability of logistics services can be achieved by services using such technologies. It will, in turn, lead to consumers’ peace of mind.
- (2) “Resilience of the supply chain”. The results of this research showed once again that logistics companies are highly dependent on and pay attention to China. Disruptions in the supply chain have continued for more than a year. Moreover, they are expected to continue for some time, even after the COVID-19 problem is resolved. In order to improve the supply chain sustainability in the future, it is necessary to consider what measures are necessary to enable rapid recovery from the disruption, including the issue of dependence on China, which is one of our future works.

References

- Bloom, N., Bunn, P., Mizen, P., Smietanka, P. and Thwaites, G. (2020), “The impact of COVID-19 on productivity (No. W28233)”, Working paper, National Bureau of Economic Research, Cambridge, MA, available at: https://www.nber.org/system/files/working_papers/w28233/revolutions/w28233.rev0.pdf.
- Brinca, P., Duarte, J.B. and Faria-e-Castro, M. (2020), “Measuring sectoral supply and demand shocks during COVID-19”, Frb St. Louis Working Paper, (2020-011).
- Choi, T.M. (2020), “Innovative ‘bring-service-near-your-home’ operations under Corona-virus (COVID-19/SARS-CoV-2) outbreak: can logistics become the messiah?”, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 140, p. 101961.
- Golan, M.S., Jernegan, L.H. and Linkov, I. (2020), “Trends and applications of resilience analytics in supply chain modeling: systematic literature review in the context of the COVID-19 pandemic”, *Environment Systems and Decisions*, Vol. 40, pp. 222-243.
- Govindan, K., Mina, H. and Alavi, B. (2020), “A decision support system for demand management in healthcare supply chains considering the epidemic outbreaks: a case study of coronavirus disease 2019 (COVID-19)”, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 138, p. 101967.
- Harris, R. (2020), “How will COVID-19 affect productivity in the UK?”, available at: <https://www.dur.ac.uk/research/news/item/?itemno=41707> (accessed 20 February 2021).
- Hendricks, K.B. and Singhal, V.R. (2003), “The effect of supply chain glitches on shareholder wealth”, *Journal of Operations Management*, Vol. 21 No. 5, pp. 501-522.
- Hinton, G. and Roweis, S.T. (2002), “Stochastic neighbor embedding”, *NIPS*, Vol. 15, pp. 833-840.
- Hosseini, P., Hosseini, P. and Broniatowski, D.A. (2020). “Content analysis of Persian/Farsi Tweets during COVID-19 pandemic in Iran using NLP”, arXiv Preprint arXiv:2005.08400.
- Ivanov, D. (2020), “Predicting the impacts of epidemic outbreaks on global supply chains: a simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case”, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 136, p. 101922.

-
- Jelodar, H., Wang, Y., Orji, R. and Huang, S. (2020), "Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: Nlp using lstm recurrent neural network approach", *IEEE Journal of Biomedical and Health Informatics*, Vol. 24 No. 10, pp. 2733-2742.
- Karmaker, C.L., Ahmed, T., Ahmed, S., Ali, S.M., Moktadir, M.A. and Kabir, G. (2021), "Improving supply chain sustainability in the context of COVID-19 pandemic in an emerging economy: exploring drivers using an integrated model", *Sustainable Production and Consumption*, Vol. 26, pp. 411-427.
- Kawasaki, T., Hanaoka, S. and Nguyen, L.X. (2015), "Inland cargo flow modelling considering shipment time variability on cross-border transport", *Transportation Planning and Technology*, Vol. 38 No. 6, pp. 664-683.
- Lin, Q., Zhao, S., Gao, D., Lou, Y., Yang, S., Musa, S., Wan, M., Cai, Y., Wang, W., Yang, L. and He, D. (2020), "A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action", *International Journal of Infectious Diseases*, Vol. 93, pp. 211-216.
- Mahajan, K. and Tomar, S. (2020), "Here today, gone tomorrow: COVID-19 and supply chain disruptions", *American Journal of Agricultural Economics*, forthcoming, available at: https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3596720.
- Mikolov, T., Chen, K., Corrado, G. and Dean, J. (2013), "Efficient estimation of word representations in vector space", arXiv Preprint arXiv:1301.3781.
- Rossi, R., Succi, V., Talevi, D., Mensi, S., Niolu, C., Pacitti, F., Di Marco, A., Rossi, A., Siracusano, A. and Di Lorenzo, G. (2020), "COVID-19 pandemic and lockdown measures impact on mental health among the general population in Italy", *Frontiers in Psychiatry*, Vol. 11, p. 790.
- SAS (2021), "Natural Language Processing (NLP) - what it is and why it matters", available at: https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html (accessed 20 February 2021).
- Sarkis, J. (2020), "Supply chain sustainability: learning from the COVID-19 pandemic", *International Journal of Operations and Production Management*, Vol. 41 No. 1, pp. 63-73.
- Serrano, J.C.M., Papakyriakopoulos, O. and Hegelich, S. (2020), "NLP-based feature extraction for the detection of COVID-19 misinformation videos on Youtube", *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*.
- Singh, B. and Singh, H.K. (2010), "Web data mining research: a survey", *2010 IEEE International Conference on Computational Intelligence and Computing Research*, pp. 1-10.
- Van der Maaten, L. and Hinton, G. (2008), "Visualizing data using t-SNE", *Journal of Machine Learning Research*, Vol. 9 No. 11, pp. 2579-2605.
- Westergaard, D., Stærfeldt, H.H., Tønsberg, C., Jensen, L.J. and Brunak, S. (2018), "A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts", *PLoS Computational Biology*, Vol. 14 No. 2, e1005962.
- Yang, L., Liu, J., Lu, Q., Riggs, A.D. and Wu, X. (2017), "SAIC: an iterative clustering approach for analysis of single cell RNA-seq data", *BMC Genomics*, Vol. 18 No. 6, p. 689.

Corresponding author

Enna Hirata can be contacted at: enna.hirata@platinum.kobe-u.ac.jp

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com