A thematic analysis–based model for identifying the impacts of natural crises on a supply chain for service integrity: a text analysis approach

Mohammad Reza Sheikhattar1 · Navid Nezafati1 · Sajjad Shokouhyar1

Received: 11 December 2021 / Accepted: 6 June 2022 / Published online: 17 June 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Numerous studies have been conducted to identify the effects of natural crises on supply chain performance. Conventional analysis methods are based on either manual filter methods or data-driven methods. The manual filter methods suffer from validation problems due to sampling limitations, and data-driven methods suffer from the nature of crisis data which are vague and complex. This study aims to present an intelligent analysis model to automatically identify the effects of natural crises such as the COVID-19 pandemic on the supply chain through metadata generated on social media. This paper presents a thematic analysis framework to extract knowledge under user steering. This framework uses a text-mining approach, including co-occurrence term analysis and knowledge map construction. As a case study to approve our proposed model, we retrieved, cleaned, and analyzed 1024 online textual reports on supply chain crises published during the COVID-19 pandemic in 2019–2021. We conducted a thematic analysis of the collected data and achieved a knowledge map on the impact of the COVID-19 crisis on the supply chain. The resultant knowledge map consists of five main areas (and related sub-areas), including (1) food retail, (2) food services, (3) manufacturing, (4) consumers, and (5) logistics. We checked and validated the analytical results with some field experts. This experiment achieved 53 crisis knowledge propositions classified from 25,272 sentences with 631,799 terms and 31,864 unique terms using just three user-system interaction steps, which shows the model’s high performance. The results lead us to conclude that the proposed model could be used effectively and efficiently as a decision support system, especially for crises in the supply chain analysis.

Keywords Supply chain crisis management · Supply chain risk monitoring · Pandemic crisis · Thematic analysis model · Text-mining

Introduction
Global supply chains (SC) have always been vulnerable to shocks, especially in exporting countries (Trautrims et al. 2020; Shanker et al. 2021). This vulnerability stems from the factors that disrupt the integration of the production of goods and services between these countries and their importer trading partners (Maillet et al. 2019). Recently, the COVID-19 pandemic has posed serious threats and risks to the SC in all sectors (Karwasra et al. 2021; Sharma et al. 2020). These risks are mainly due to lockdown measures that governments adopt and apply in the form of health policies against the COVID-19 outbreak, which imposes restrictions on “movement,” “production,” “logistics,” and “business activities” (Menut et al. 2020; Karwasra et al. 2021).

For example, China was one of India’s primary importers of raw materials for drug production between 2018 and 2019. Due to the pandemic, the import of raw materials was disrupted, and the Indian pharmaceutical industry faced shortages and delays in the supply and distribution of pharmaceutical products (Guerin et al. 2020). The shipping industry, which involves about 90% of world trade, has been severely affected by government lockdown policies. Hence,
the World Trade Organization (WTO) announced that the volume of world trade has decreased by about 23% by the end of 2020 (Palme et al. 2022). Also, due to disruption in the food SC, 285 million people worldwide suffered from severe starvation by the end of 2020 (Rizou et al. 2020).

Such significant changes significantly increase the need to examine the pandemic effects on the SC as an essential and distinctive case study. In this regard, due to the extensive discussions and scientific research that is conducted on the consequences of the COVID-19 crisis on social media, this media has become a constructive database for identifying and analyzing the risk in the SC (Aghion et al. 2008; Bakhtin et al. 2020; Spitsberg et al. 2013; Momeni and Rost 2016).

Various quantitative and qualitative methods have been used to analyze the SC crisis in the literature (Wan et al. 2021). The comprehensiveness of the findings in qualitative methods is questionable due to the limitations of the samples (Ponis and Ntalla, 2016). There is also concern about the bias of results in these methods due to human judgments (Arlinghaus et al. 2020; Schorsch et al. 2017; Puljić, 2010).

Following the increase in data and analytical news on social media, research has recently moved to the use of quantitative data-driven methods. However, these methods also have major drawbacks. Due to the nature of risk, which is ambiguous and complex and has different meanings, processing these data by pure data mining methods is impossible and requires human insight to identify risks.

This scientific gap in the analysis and identification of crisis management continues. Therefore, one of the serious motivations for this research is to provide a practical framework for analyzing and identifying the effects of a crisis on the SC using a large amount of available data. This framework gathers the contents of the online analytical reports, converts these data into interpretable structured information, and finally provides classified knowledge on crisis analysis through user interactions. The framework could be used in decision support systems of companies or governments. To validate the proposed model’s effectiveness, we collected 975 online reports about the impact of COVID-19 on SCs between 2019 and 2021 and analyzed them using the framework.

In general, this paper presents a framework to analyze the effects of a crisis on the SC and reports the results from a case study to evaluate the effectiveness of the proposed framework. The results user achieves from interactions using thematic analysis could also be used to unfold the research gap and future research guidelines in SC crisis analysis.

The rest of the paper is organized as follows. The “Literature review and related works” section provides available knowledge for the research topics and related research and expresses the research gap. The “Methodology: framework for identifying the effects of natural crises on the supply chain” section presents the methodology used in this research. The “Case analysis” section evaluates the framework’s effectiveness by analyzing the impact of COVID-19 on the SC case study. The “Discussion” section discusses the research findings, and finally, the “Conclusion” section concludes the paper.

Literature review and related works

Text analysis technical methods

Various text analysis techniques related to this research are discussed in this section, including Word2vec, knowledge map, TFIDF, topic modeling, and text-mining.

Word2vec model

In order to process texts, words must be converted to numerical quantities. One of the most widely used methods is the word2vec model. The Word2vec model is a neural network–based model that converts words into numeric vectors (Mikolov et al. 2013). Its input is a text corpus, and its output is a set of real-valued feature vectors; each vector represents or encodes the meaning of one word in the corpus. Thus, the words closer in vector space are expected to be similar in meaning (Jurafsky and Martin 2008). There are two models, Continuous Bag-of-Words (CBOW) and Skip-gram, for learning underlying word representations for each word by using neural networks in the word2vec model. The CBOW model is used for word2vec in this paper.

Co-occurrence network and knowledge map

One of the text analysis methods used in this research is the co-occurrence network, also referred to as the knowledge map (KM) (Segev 2020). This method involves visualizing potential relationships between concepts or other entities represented in textual data (Freilich et al. 2010). The process of building a co-occurrence network includes (Segev 2021):

- Identifying keywords in the text,
- Calculating the frequencies of terms, and
- Analyzing networks for finding central terms and clusters of terms in the network.

In co-occurrence networks, the connecting pairs of terms are defined based on their paired presence in a textual unit. Co-occurrence networks are generated by linking pairs of terms using a set of criteria that define concurrency (Freilich et al. 2010).
TFIDF

The term frequency-inverse document frequency (TFIDF) is determined to filter public words from the dataset and highlight the most important words (Jing et al. 2002). It has been widely used to improve processing accuracy as an indicator of general and important terms (Gudivada et al. 2018). The principle of applying TFIDF is explained as follows. By applying this method, terms that are frequently used in any document (such as conjunction words) have a low rank (Qaiser and Ali 2018). The equations (Eqs. 1–3) show TFIDF calculation:

\[tf(t, d) = \frac{f_{t,d}}{\sum_{r \in d} f_{r,d}}\]  

(1)

\[idf(t, D) = \log(N \cdot (1 + |\{d \in D : t \in d\}|))\]  

(2)

\[TFIDF(t, d, D) = tf(t, d) \cdot idf(t, D)\]  

(3)

In the above equations, \(f_{t,d}\) is the number of times that term \(t\) occurs in document \(d\), \(N\) is the number of documents in the corpus, and \(|\{d \in D : t \in d\}|\) is the number of documents where the term \(t\) appears. If the term is not in the corpus, this will lead to a division-by-zero. Therefore, it is common to adjust the denominator to \(1 + |\{d \in D : t \in d\}|\).

**Topic modeling**

In machine learning language and natural language processing, topic modeling is a distinctive statistical model for discovering essential topics in a set of documents. Topic modeling is a widely used text-mining tool to find hidden semantic structures in a text (Blei 2012). Since each document is about a specific topic, certain words are expected to appear in its textual content. The topics extracted from topic modeling technique are a group of similar words cluster. Topic models are statistical algorithms to discover the hidden semantic structures of a text (Cao and Li 2007). One of the most common methods for topic modeling is the latent Dirichlet allocation (LDA) (Blei 2012). LDA is an unsupervised learning model used as a topic modeling technique that can classify text in a document into specific topics. This technique uses the Dirichlet distribution to discover topics for each document model and words for each topic model.

Figure 1 presents the core components of LDA algorithm, where \(K\) is the number of topics, \(N\) is the number of words in the document, \(M\) denotes the number of documents, \(\alpha\) is the parameter of the Dirichlet prior to the per-document topic distributions, \(\beta\) is the same parameter of the per-topic word distribution, \(\phi(k)\) is the word distribution for topic \(k\), \(\theta(i)\) is the topic distribution for document \(i\), and \(Z(i, j)\) is the topic for the \(j\)th word in document \(i\). Equation 4 is shown below:

\[p(w, z + \theta, \phi|\alpha, \beta) = \prod_{i=1}^{N}p(\theta|\alpha)p(z|\theta)p(\phi|\beta)p(w, z, \theta)\]  

(4)

Each word in the corpus is randomly classified into a subject and tagged with a specific subject number. Then, according to the algorithm sample function, a new subject number is assigned to each word. This process continues until it converges.

**Text-mining and its applications in supply chain management**

The existing literature confirms the critical role of artificial intelligence in various industry fields such as healthcare, education, crisis management, and production. For example, Sharifi et al. (2021) studied the impact of artificial intelligence and digital style on industry and energy after the COVID-19 pandemic. Their object was to investigate the effects of COVID-19 on the various fields of medicine, industry, and energy. Artificial intelligence technology tries to improve the efficiency of the management process during the crisis response. In particular, they have studied the effect of artificial intelligence and digital style in reducing the damage of this deadly virus. Ahmadi et al. (2021) used the capabilities of artificial intelligence to provide an extended pandemic model for estimating the COVID-19 epidemic and assessing its risks. They presented a generalized logistics growth model (LGM) to estimate COVID-19 outbreak sub-waves in Iran. Nasirpour et al. (2021) used multivariate spatial autoregressive (MSAR) to investigate the relationship between solar activity and COVID-19 and to predict possible future viruses.

Artificial intelligence technology tries to improve the efficiency of the management process during the crisis response. Baryannis et al. (Baryannis et al. 2019) conducted a comprehensive review of the SC literature that addresses problems relevant to SCRM using approaches within the range of artificial intelligence. To this end, they conducted
research on different definitions and classifications of SC risk and related concepts such as uncertainty. Then, a mapping study was performed to categorize the existing literature based on the AI method used, from mathematical programming to machine learning and big data analysis, and the specific SCRM task they address. Their research points out that while risk management is fraught with challenges, identifying sources of risk-related information is one of the primary concerns.

This type of data shows the importance of text-mining, a subset of artificial intelligence in analyzing SC risk data. Text-mining, also known as text data mining, transforms unstructured text into a structured format to identify meaningful patterns and new insights. In text-mining, descriptive and predictive analyzes can be used (Dang and Ahmad 2014). Typical text-mining tasks include the following:

- Text categorization and clustering,
- Concept/entity extraction,
- Association rule mining,
- Sentiment analysis,
- Document summarization,
- Visualization,
- Entity-relation modeling (i.e., learning relations between named entities).

For example, Chiu and Lin (2018) merged text-mining and Kansei Engineering (Nagamachi 1995) to derive Kansei’s descriptive terms based on actual customer surveys and use it to predict the design of a consumer-preferred product while reducing the specific repetitive tasks of designers. The accuracy of traditional text-mining in the analysis of texts, especially texts with complex meanings, is very low because it cannot use the semantic information of the text effectively (Hu and Zhang 2010).

Ontology can extract key concepts and inter-relation and thus provide a common understanding of a domain. Combining these two techniques can obtain accurate text-mining analysis (Hu and Zhang 2010). Elbattah et al. combined domain ontology and text analytics techniques to analyze data in healthcare (Elbattah et al. 2021). Text-mining is also used to access knowledge about patents, known as patent knowledge retrievals. Liu et al. proposed a function-based patent knowledge retrieval tool for the conceptual design of innovative products.

These previous studies show that the emerging field of text-mining can turn natural language into practical results, gain new insights, manage information loads, and use artificial intelligence in decision-making. One research area in which text-mining has been widely used is the SC. The advent of modern information technologies such as IoT, big data, blockchain, and artificial intelligence have created new opportunities for efficient SC management. For example, Akundi et al. gathered information from various textual sources (e.g., tweets, news, and other social media) to understand how textual data about a given smartphone could affect its SC and management (Akundi et al. 2018). Mayer et al. examined how text-mining could provide insights into the impact of the coronavirus epidemic on SCs, focusing on epidemic consequences for SC structures related to risk, flexibility, and sustainability (Dowling et al. 2019). They showed that some SC topics, such as risk, flexibility, disruption, and consistency, differ in their news coverage on the type of newspaper and the number of coronavirus disease 2019 (COVID-19) infections.

Aday and Aday (2020) evaluated the impact of COVID-19 on the agri-food sector. They summarized the recommendations needed to reduce and control the effects of the pandemic by using textual report analysis. Su and Chen (2018) developed a Twitter-enabled supplier status assessment tool to improve supplier selection. They applied text-mining on Twitter tweets to retrieve supplier-related information and analyze potential risks and uncertainty. They used text-mining in Twitter tweets to retrieve supplier information and analyze risks and potential uncertainties. They applied text-mining on Twitter tweets to retrieve supplier-related information and analyze potential risks and uncertainty. The proposed method was proven to improve the efficiency and accuracy of cross-border e-commerce (CBEC) commodity risk evaluation.

Research gap

Numerous qualitative methods have been developed to analyze the effects of the crisis on the SC and related risks, such as fuzzy cognitive maps (Bevilacqua et al. 2020), fuzzy AHP approach (Nazam et al. 2020), pattern matching technique (Köksal et al. 2018), analytic network process (Martino et al. 2017), explorative qualitative study (Kam et al. 2011), qualitative survey approach (Moon et al. 2010), and Delphi method (Cerruti and Delbufalo 2009). Many of these methods have validation challenges due to limited samples (Ponis and Ntalla, 2016). On the other hand, qualitative methods based on human judgments have raised concerns about possible biases (Giannakis and Papadopoulos 2016).

Today, large amounts of SC crisis data are available as text-based information from social networks, open portals, and databases for academic purposes. In this regard, new research has been directed toward data-driven methods, a text analysis approach to monitoring SC crises and understanding risk patterns (Yan et al. 2019; Chu et al. 2019; Shah et al. 2021; Kara et al. 2020; Da Silva et al. 2020).

However, due to the nature of risk data, which are complex, ambiguous, and contradictory, pure data mining methods do not meet the required precision, effectiveness, and efficiency; and even may mislead (Wang and Ye 2018). In
In this regard, using the ability of text-mining algorithms to analyze a large amount of data with assist of human intelligence to guide the algorithm and iteratively modify the results can fill the gap in identifying risk factors and investigating risk effects.

**Methodology: framework for identifying the effects of natural crises on the supply chain**

Our proposed framework is based on the thematic analysis combined with text-mining methods (Braun and Clarke 2006; Guest et al. 2012). According to Fig. 2, the framework consists of three main layers:

- “Data gathering” on related topics;
- “Pre-processing” to reduce noise and data preparation;
- “Knowledge discovery” to analyze and extract knowledge in an interaction with the user.

The layers are explained in the following subsections.

**Data gathering layer**

The data gathering layer collects data from data sources, generally web pages and social media posts. This layer contains one component for data collection and delivers the collected data to the next layer, pre-processing layer. Although data sources are in different formats, e.g., text, image, voice, and video, we focus on the text format because natural language processing techniques are more mature than processing other formats. Since only the text sources are considered in the proposed platform, this layer prepares the data in the form of a text corpus for the pre-processing layer.

This research focused on identifying the impacts of natural crises on the food SC. Therefore, the data sources include newsletters, reports, and events on food, risks, logistics, freight, operations, regulations, and technology.

**Pre-processing layer**

In Fig. 2, the pre-processing layer receives the corpus from the data gathering layer. It prepares some transformed corpus forms for other processes in the knowledge discovery layer. As shown in the pseudo-code of Fig. 3, this layer contains four components to perform four actions on the corpus and prepare transferred forms of the corpus, correspondingly:

- Normalized_corpus creation: This component removes all URLs, stop words (e.g., “a,” “an,” “them,” “it”), and special characters (e.g., HTML and XML tags) from the text. Furthermore, this component replaces lemmatized form of every word; therefore, the words with the same base also have the same form.
Algorithm: Pseudo-code for the Pre-processing

1: \textbf{Input:}\ corpus [as a global variable]
2: \text{normalized_corpus} = \text{NORMALIZE_CORPUS}(\text{corpus}) [Remove URLs, stop words, special characters, and lemmatization]
3: \text{tokenized_corpus} = \text{TOKENIZE_CORPUS}(\text{corpus}) [tokenize all corpus content]
4: \text{word2vec_model} = \text{WORD2VEC_MODEL_GENERATION}(\text{tokenized_corpus})
5: \text{corpus_sentences} = \text{SENTENCE_SPLITTING}(\text{corpus}) [split corpus to its constitutive sentences]

Fig. 3 Pseudo-code for the Pre-processing

- Tokenized_corpus creation: This component splits the input corpus into its constituent tokens, chunks of information that can be considered discrete elements with a useful semantic unit for processing.
- Word2Vec Model generation: This component generates the Word2Vec model of the corpus, which is used in the knowledge discovery layer to identify similar terms.
- Corpus_sentences creation: Sentences are considered knowledge units (KUs) in the proposed framework. Therefore, this component divides the corpus into separate sentences.

Knowledge discovery layer

The core part of the framework is the knowledge discovery layer, which provides a user-system interaction facility to extract knowledge iteratively. As shown in Fig. 2, this layer consists of three component groups: crisis KM handling, crisis query handling, and knowledge extraction/evaluation. Figure 4 shows the corresponding pseudo-code of this layer. More details are described below.

Crisis knowledge map handling

As discussed in the “Literature review and related works” section, KM is represented as a network whose nodes are terms from the tokenized corpus, and the links are the meaningful relations between the nodes. The KM is created based on a top-down network of terms. The user could select any term from any level to be expanded into more related terms at the next level as co-occurrence terms. This component group comprises LDA_Cluster Visualization, Expand_KM, and Visualize_KM.

LDA_Cluster visualization To create a KM, it is necessary to extract important topics in the whole corpus as key terms. These key terms, which form the first level of the KM, are selected by the user with the assistance of topic modeling and visualization of clusters. At the first step, the KM is initialized to a single node, root_node (line 5 of Fig. 4). Then, clusters are identified using LDA (line 6 of Fig. 4) and visualized as word clouds (lines 7–9 of Fig. 4). Afterward, level 1 of KM is constructed, adding the terms selected by the user from the visualized clusters to the initialized KM with the associated links to the root_node (lines 10–14 of Fig. 4). Level 1 of KM is shown to the user for further expansion of the KM (line 15 of Fig. 4).

Expand_KM After creating the first level of KM, in this step, each selected term (which is called key_term) expands to a subset of relevant terms, called co-occurrence terms, at lower levels in KM. This operation is repeated for each selected term in each layer and expands to lower layers based on user preferences until the user stops the expansion operation. This is shown in the pseudo-code of Fig. 3. The main part of the program is the iterative loop, located on lines 16–26. The algorithm then uses the Expand_KM procedure to expand this expression to co-occurrence terms at the lower layer. The operation of Expand_KM component is shown as the Expand_KM procedure (lines 28–36 of Fig. 4).

The Expand_KM procedure performs three main actions: (1) extracting context terms, (2) filtering unnecessary terms, and (3) extending the knowledge map.

Extracting context terms At first, it extracts context terms using Extract_Context_Terms function, in which the context_terms and the number of their occurrences are extracted. The context terms are terms in the neighborhood of the central selected key_term in the win_size range throughout the tokenized_corpus. The context terms and their relationship to the key term and the window size range are shown in Fig. 5.

Filtering unnecessary terms As the second action, Expand_KM filters unnecessary terms. Therefore, the co-occurrence_terms are extracted as a key_term expansion in the KM. These co-occurrence terms are a subset of context terms that have a meaningful relationship with the key term and have not been randomly placed in the neighbor of the key term. In order to extract co-occurrence terms, two filtering operations are applied to context_terms to remove unnecessary terms: based on the TFIDF threshold and based on the phi threshold.

The TFIDF threshold filter (line 30 of Fig. 4) calculates the TFIDF parameter of the context_terms obtained from the previous step and removes terms whose TFIDF values
Algorithm: Pseud-code for the Knowledge Discovery

1: Input: corpus, normalized_corpus, corpus_sentences [As global variables]
2: Global variables:
3: $T_{TFIDF}, T_{Phrase}, T_{phi}$ [Thresholds for TFIDF, Frequency Occurrence, and phi]
4: win_size [window size]
5: KM.node = [root_node], KM.links = Null [Initialize Knowledge Map (KM) to one root node and no link]
6: $C_1, …, C_k$ = LDA_CLUSTER(normalized_corpus) [Topic modeling using LDA to cluster corpus to k topic collections, $C_1, …, C_k$]
7: For $i$ from 1 to $k$ do [for each cluster visualize the cluster]
8:    WORD_CLOUD($C_i$) [for words in cluster $C_i$]
9: EndFor
10: For $i$ from 1 to $k$ do [Create KM Level 1 as KM initialization]
11:    $t = GET_KEY_TERM()$ [Get a term from user as the key term, considering the word cloud]
12:    add $t$ to KM.nodes [each node is a term]
13:    add link <root_node, $t$> to KM.links
14: EndFor
15: VISUALIZE_KM() [Visualize knowledge map, KM, for user]
16: Repeat [KM expansion according to user’s preferences]
17:    key_term = GET_KEY_TERM() [User selects a term (or node) in KM]
18:    $T_{TFIDF}, T_{Phrase}, T_{win_size}$ = GET_THRESH_WIN_SIZE() [User can tune any threshold and/or window size]
19:    EXPAND_KM(key_term) [Expand Knowledge Map (corresponding to the Expand_KM component in Figure 2)]
20:    VISUALIZE_KM() [Visualize knowledge map, KM, for user (corresponding to the Visualize_KM component in Figure 2)]
21:    query = GET_QUERY [query is a traverse of terms from root_node to a leaf node in KM selected by user]
22:    enriched_query = ENRICH_QUERY(query) [propose some semantically similar terms to the terms in query, and user decides adding some of them to query and make enriched_query]
23:    ku_array = EXTRACT_KU(enriched_query, corpus_sentences) [a list of sentences (knowledge units, KUs) are extracted from corpus_sentences and sorted based on their similarity to enriched_query]
24:    evaluated_KU_array = EVALUATE_KU(ku_array) [The KUs are evaluated and sorted based on “category information value”, then user selects the most valuable sentences.]
25:    user_flag = USER_CONFIRM() [get user choice whether satisfied (True) or not (False)]
26: Until user_flag
27: PRINT evaluated_KU_array [The final findings of KUs (sentences) as the result of user-system interactions]

28: Procedure EXPAND_KM(key_term)
29:    context_terms = EXTRACT_CONTEXT_TERMS(Tokenized_corpus, key_term, win_size) [Extract context terms in the range of win_size around every occurrence of key_term in the tokenized_corpus (as Figure 5)]
30:    imp_terms = FILTER(context_terms, $T_{TFIDF}$) [extract important terms from context_terms such that (terms’ TFIDF)$>T_{TFIDF}$]
31:    co_occur_terms = FILTER(imp_terms, $T_{phi}$) [extract co-occurrence terms from imp_terms such that (terms’ phi)$>T_{phi}$]
32: For each term $t$ in co_occur_terms do
33:    add $t$ to KM.nodes [each node is a term]
34:    add link <key_term, $t$> to KM.links
35: EndFor
36: EndProcedure

37: Function ENRICH_QUERY(query) [Query is a sequence of terms]
38: Initialize positive_terms and negative_terms
39: For each term in query
40:    choice = GET_USER_CHOICE() [User specifies term to be in positive_terms set or in negative_terms set]
41:    If choice is positive Then
42:        add term to positive_terms
43:    Else [choice is negative]
44:        add term to negative_terms
45: EndIF
46: EndFor
47: Similar_Term $\leftarrow$ EXTRACT_SIMILAR_TERM (positive_terms - negative_terms) [based on Word2Vec technique that finds most similar terms to (positive_terms - negative_terms) terms]
48: selected_Similar_Term = GET_SELECTED(Similar_Term) [user selects some terms to selected_Similar_Term]
49: enriched_query = positive_terms $\ast$ selected_Similar_Term
50: Return enriched_query
51: EndFunction

Fig. 4 Pseudo-code for the knowledge discovery layer
are less than the TTFIDF threshold tuned by the user (line 18 of Fig. 4). After filtering operation in this step, the most important terms from freq_terms are extracted and stored in the imp_terms (important terms) variable.

The second filter, the phi filter, is based on the phi coefficient threshold. The phi coefficient is used to measure the correlation of important terms with the key term. As shown in line 31 of Fig. 4, in this filter, imp_terms obtained from the previous step whose \( \phi \) values are not within the phi threshold \( (T_{phi}) \) are removed in this step. The remaining terms of the filtering operation are stored in the co_occur_terms variable. Table 1 shows the required values for calculating the \( \phi \) parameter of words \( X \) and \( Y \) in a set of documents.

In Table 1, \( n_{11} \) represents the number of documents that contain both the words \( X \) and \( Y \), \( n_{00} \) represents the number of documents that contains none of them, and \( n_{10} \) equals the number of documents containing \( X \) and do not contain \( Y \), and similarly, \( n_{01} \) equals the number of documents containing \( Y \) and does not contain \( X \). The total values are also shown in the table, \( n \) equals the number of all documents. Equation 5 shows the calculation of the coefficient \( \phi \) to obtain the correlation between the words \( X \) and \( Y \) in Table 1.

\[
\phi = \frac{(n_{11}n_{00} - n_{10}n_{01})}{\sqrt{n_{11}n_{00}n_{01}n_{10}}} \tag{5}
\]

The correlation coefficient \( \phi \) varies from \(-1\) to \(+1\). The value of \(+1\) indicates the maximum agreement, the value of \(-1\) indicates the disagreement, and the value of \(0\) shows no relationship.

**Extending the knowledge map** As the third action of \textsc{Expand\_KM}, all the terms are considered as nodes, and also their relation with the key term from which they are expanded is stored to create and visualize a KM. Each term in the co_occur_terms variable is added to the \textsc{KM\_nodes} as a node, and the relation between this node and the key term from which it is generated is stored in the \textsc{KM\_links} variable (lines 32–35 of Fig. 4).

**Visualize\_KM** This component is used to visualize the relationship between key terms and their co-occurrence terms. This operation is shown in the \textsc{Visualize\_KM} component in Fig. 2. After expanding the knowledge map in the previous step, the nodes and the relationships between the nodes are saved in \textsc{KM\_variable} (containing \textsc{KM\_nodes} for nodes and \textsc{KM\_links} for links). The \textsc{Visualize\_KM} function visualizes the knowledge map for the user (line 20 of Fig. 4).

**Crisis query handling**

**Enrich\_Query** To extract the relevant KUs, it is necessary to have a proper query. To create a suitable query, finding the correlated terms in the relevant knowledge domain is essential. The KM created in the previous step expresses the relationship between terms in a (tree-like) directed
acyclic graph. The query is a traverse of terms from root_node to a leaf node in the KM network selected by the user. The request extracted from KM may not include all aspects, and it may be necessary to add or remove concepts. For this purpose, the query is sent to the Enrich_Query component for enrichment (Fig. 2). This operation is presented by the pseudo-code of the simulation in Fig. 4. In line 21 of the pseudo-code by the Get_Query function, the algorithm gets the query that the user has created from the KM and puts it in the query variable. In line 22 of the pseudo-code, the Enrich_Query function creates an enriched query. Details of the Enrich_Query function are shown in lines 37–47 of the pseudo-code. Due to the algebraic ability in word embedding to add and remove terms, we used “+” and “−” signs to add and remove concepts (terms) to create a suitable query. Two variables, positive_terms and negative_terms, are initialized (line 38). The terms that the user puts in the positive_terms variable are the concepts to be added, and the terms he puts in the negative_terms variable are those to be removed. In lines 39 to 45 of the pseudo-code, the user decides which terms to be added to each variable. The related mathematical operation is:

\[
q_i' = \arg\max_{q \in V} (q \cdot (q_p - q_n))
\]

(6)

where \(q_p\) represents the “sum of positive terms vector” and \(q_n\) represents the “sum of negative terms vectors.” The system extracts \(q_i'\) for enriching query from the vocabulary \(V\) (the set of unique words used in the text corpus). This operation is performed by finding the term with maximum angular similarity to the vocabulary vector (expressed as vector dot product, assuming all term vectors have the same length).

The user selects some terms in the Similar_Term variable as selected_Similar_Term to enrich the query, as shown in line 48 of pseudo-code of Fig. 4. To create an enriched query, selected_Similar_Term is added to positive_terms to create an enriched_query variable, as shown in line 49 of pseudo-code in Fig. 4. As a result, mathematically \(q_i\) (enriched query vectors) are created according to Eq. 7. The query enrichment operation is shown in the framework presented in Fig. 6.

\[q_i = q_p + q_i'\]

(7)

Knowledge extraction/evaluation

Extract_KU After creating a user-interested crisis query, the Extract_KU component of Fig. 2 extracts the most similar KUs to the query. KUs are the most similar corpus sentences to enriched queries representing knowledge propositions. The KUs extracted by the machine allow the analyst to make a hypothesis about the crisis. It also steers the analyst’s subsequent searches to select appropriate key terms. It is necessary to convert text words into numerical quantities to conduct calculations on queries and KUs. We have used the word2vec algorithm, which processes text by vectorizing the words.
Each SC document \( (SC_i; i = 1, 2, \ldots, m) \) consists of \( n \) numbers of terms \( t_1, t_2, t_3, \ldots, t_n \) representing the spatial–temporal information of an SC. The basic idea is to extract unique content-bearing terms from the set of SC documents and afterward assign weight to every term based on the product of term frequency and inverse document frequency (TFIDF). Consequently, these terms should be treated as a numerical representation of the features of the similarity algorithm. Therefore, we have represented an SC document \( (SC_i) \) through the \( d \)-dimensional feature vector in the term space as:

\[
SC_i = W_1 t_1, W_2 t_2, W_3 t_3, \ldots, W_n t_n,
\]

where \( W_k \) is the weight assigned to each term \( t_j \) in the SC document as:

\[
W_n = \text{F}rq(t_j, SC_i) \ast \text{IDF}
\] (8)

where \( \text{F}rq(t_j, SC_i) \) is the frequency of the term \( j \) in a SC document \( i \) and IDF is the inverse document frequency calculated as:

\[
\text{IDF} = 1 + \log \left( \frac{\text{Number of SC documents containing the } t_j}{\text{Total number of SC documents}} \right)
\] (9)

The numerical representation of the SC sentences vector (KUs vector) was thus represented through these weighted SC terms. Equation 10 shows the equivalent formal mathematical description for sentence vector:

\[
KUVector = \sum_{i=1}^{m} \left( \text{F}rq(t_j, SC_i) \ast \text{IDF} \ast \left( \frac{TV_{t_j} : t_j \in WV}{m + \sigma} \right) \right)
\] (10)

where \( SC_i \) refers to each textual document in the corpus. \( TV_{t_j} \) is the term vector of each term \( t_j \). \( \text{F}rq(t_j, SC_i) \) is the frequency of \( t_j \) in each SC \( (SC_i) \). \( \text{IDF} \) is the reverse frequency of \( t_j \) in each SC \( (SC_i) \). \( m \) is the total number of terms in each sentence. A fixed value \( \sigma \) has been added to prevent the denominator from becoming zero.

However, we used cosine similarity as the distance function in the similarity calculation between queries and KUs according to Eq. 11, and then KUs will be ranked accordingly.

\[
\text{Similarity}(\bar{q}, KU_j) = \frac{\bar{q} \cdot KU_j}{\|\bar{q}\| \ast \|KU_j\|}
\] (11)

Figure 4 shows the simulation code of the above operation. In line 23, The \text{Extract} _\text{KU} function takes the two parameters \text{enriched} _\text{query} and \text{corpus} _\text{sentences}, extracts a list of sentences (KUs) from \text{corpus} _\text{sentences}, and ranks them based on their similarity \text{enriched} _\text{query}.

The \text{Extract} _\text{KU} function takes two parameters, \text{enriched} _\text{query}, and \text{corpus} _\text{sentences}, extracts a list of the most similar sentences in \text{corpus} _\text{sentences} to \text{enriched} _\text{query} and ranks them by their similarity values.

\text{Evaluate} _\text{KU} After extracting the crisis KUs, it is necessary to evaluate the information value of these KUs to create a hypothesis or knowledge with high validity. To determine the information value of the extracted KUs, first, a similarity diagram is created between the query and each crisis KU, as
shown in Fig. 7. As can be seen in this diagram, the gradient of the curve is severely broken in some places. Each break in the gradient could be interpreted as the border between two separate clusters (C1 to C5 in Fig. 7). Therefore, for each cluster \( C_j \), the information value of cluster \( c \), \( IV(c_j) \), is measured by Eq. 12.

\[
IV(c_j) = \left( \sum_{i=1}^{m} X_i^{C_j} / m \right) / \left( m * \text{Dev} \left( X_i^{C_j} \right) + m \right) \tag{12}
\]

In the above equation, \( X_i^{C_j} \) is the similarity value of \( i \)-th KU from \( j \)-th cluster. \( \text{Dev} \left( X_i^{C_j} \right) \) is the standard deviation of the \( X_i^{C_j} \). \( M \) is the number of KUs in the cluster. In line 24 of pseudo_code in Fig. 4, by EVALUATE_KU function, the KUs are evaluated and sorted based on “cluster information value” and the most valuable sentences are selected.

**Case analysis**

In this section, to evaluate the framework operation, we have implemented it and ran a case to analyze the impact of COVID-19 on the SC.

**Data collection and pre-processing**

In this case study, we collected a large unstructured volume of SC crisis reports\(^1\) from professional websites, including Supply Chain 24/7,\(^2\) Supply Chain Quarterly,\(^3\) Supply Chain Dive,\(^4\) Supply Chain Management Review,\(^5\) SupplyChainDigest,\(^6\) Supply Chain Digital,\(^7\) SupplyChainDigest,\(^8\) and Commercial Risk online\(^9\) by using the crawling technique. These websites include analytical reports and articles focusing on the supply chain in logistics, transportation, operations, logistics, regulation, technology, and risk. The topics of the articles and the titles of the interviews prepared by analysts and managers of various industries show that these information sources can be suitable for the input to the presented framework. Many of the mentioned news websites have tagged reports on numerous topics in the field of SC crisis, which significantly have facilitated access to the information we are interested in for this case. Through these websites, terms such as “supply chain crisis,” “supply chain risk AND management,” “supply chain crisis AND COVID-19 impact,” and “supply chain AND pandemic impact” were searched between the years 2019 and 2022. As a result, 1024 analytical reports on the supply chain crisis were collected. Severe inspections were carried out to remove repeated and irrelevant documentation. Finally, 975 articles have been determined to be suitable for further analysis. Then, we cleaned up and pre-processed these data to remove irrelevant information and prepare them for analysis. After the normalization operation, 30 commonly used terms in the SC corpus were extracted, shown in Table 2. The resultant corpus contains 975 reports with 631,799 terms and 31,864 unique term forms. The most effective and meaningful terms in the corpus have been shown in Table 2.

| Ranking | Term             | Frequency | Ranking | Terms   | Frequency |
|---------|------------------|-----------|---------|---------|-----------|
| 1       | Risk             | 1155      | 16      | Countries | 1652      |
| 2       | Supply chain     | 10,986    | 17      | Market   | 1546      |
| 3       | Management       | 5553      | 18      | Company  | 1493      |
| 4       | Crisis           | 3465      | 19      | Industry | 1377      |
| 5       | Pandemic         | 3015      | 20      | Demand   | 1252      |
| 6       | Companies        | 2525      | 21      | Products | 1186      |
| 7       | COVID            | 2361      | 22      | Production | 1105     |
| 8       | Business         | 2167      | 23      | Resilience | 1096   |
| 9       | Blockchain       | 2055      | 24      | Logistics | 1076      |
| 10      | Food             | 2011      | 25      | Market   | 1054      |
| 11      | Technology       | 2005      | 26      | Security | 1046      |
| 12      | Product          | 1971      | 27      | System   | 1035      |
| 13      | Disruptions      | 1823      | 28      | Suppliers | 1027      |
| 14      | Service          | 1765      | 29      | Customers | 1021      |
| 15      | Impact           | 1701      | 30      | Trade    | 1012      |

**LDA_cluster visualization**

We applied topic modeling to the analytical reports using the LDA_Cluster, and the reports were divided into five clusters with different topics. Figure 8 shows the word clouds of the clusters, highlighting the most important terms in each cluster. One of the important terms of each cluster was selected by the user to create the first layer of KM. These selected terms are as follows: “food retail,” “food service,” “manufacturing,” “consumer,” and “logistics,” which are used to create the first layer of the KM.

**Knowledge map expansion**

After constructing the first layer of the KM, the user could select any of the terms of the first layer as the key term for KM expansion. First, context_terms were extracted

---

\(^1\) Available from [https://drive.google.com/drive/folders/1tOQXmwYBR_b1Rzhl4vtFzwCpN7Y29Kcn?usp=sharing](https://drive.google.com/drive/folders/1tOQXmwYBR_b1Rzhl4vtFzwCpN7Y29Kcn?usp=sharing)

\(^2\) [https://www.supplychain247.com](https://www.supplychain247.com)

\(^3\) [https://www.supplychainquarterly.com](https://www.supplychainquarterly.com)

\(^4\) [https://www.supplychindive.com](https://www.supplychindive.com)

\(^5\) [https://www.scmr.com](https://www.scmr.com)

\(^6\) [https://www.scdigest.com](https://www.scdigest.com)

\(^7\) [https://supplychaindigital.com](https://supplychaindigital.com)

\(^8\) [https://www.scdigest.com](https://www.scdigest.com)

\(^9\) [https://www.commercialriskonline.com](https://www.commercialriskonline.com)
throughout the corpus in the win_size range. In this case study, we set the win_size to 20 for all steps; however, the user can assign any other values to it based on his insight. In the next step, two filtering operations were applied to all extracted context_terms:

- **TFIDF filtering**: The TFIDF of each context term was calculated and normalized to the 0–1 range. The terms whose TFIDF was less than the threshold value of 0.65 (set by the user) were ignored. The user determines the TFIDF threshold based on the field of study and context of analysis. The important context_terms for the selected key term were obtained at the end of this step.

- **Phi filtering**: Table 3 shows the values of parameter $\phi$ calculated for the most important context terms. The boldfaced values identify the terms with negative $\phi$ values, which are the terms that do not have a logical relationship with the key terms and should be ignored.

At the end of the filtering operation, the co-occurrence terms of a key term selected by the user were obtained. These co-occurrence terms are the most important terms that semantically describe the key term (about the impact of COVID on the supply chain) concerned by the user. For example, as Table 3 shows, the terms “shopping,” “users,” “law,” “spending,” “behavior,” “respondent,” “enforcement,” “food,” “eating” and “crisis” were extracted as the context of “consumer” key term after filtering step.

**Visualize_KM**

In this component, the relationship between the key term and the context_terms was displayed in a KM. As shown in Fig. 9, all selected terms are derived from topic modeling (e.g., “food retail,” “food service,” “manufacturing,” “consumer,” and “logistics”) at level 1 of the KM were considered. Also, the resulting context_terms at the lower level are linked to the upper key term, respectively. For example, according to the KM in Fig. 9, the terms “products” and “supply” are the underlying terms for the key term “demand,” which are attached to it in the KM.

**Enrich_Query**

Table 4 shows some details on query enrichment in the case we ran. The user selected the query “food retail + online” considering the KM to discover knowledge propositions throughout the corpus that define the relationship between “food retail” and “online” in times of crisis. For this purpose, terms related to this query were proposed by the Enrich_Query component to extract more relevant KUs. In the Enrich_Query component, the analogy rate of all terms in the corpus with the initial query was calculated using word embedding. The analogy rate of the terms “channels,” “business,” “impact,” “COVID,” “consumer,” “shopping,” “frequency,” “performer,” and “company” to “food retail” was higher than others. Therefore, these terms were suggested as terms that could enrich the query. Three terms “channels,” “impact,” and “consumer,” shown in boldface, were selected from the terms suggested by the query enrichment component. Similar operations were performed for “food retail + worker,” “food service + sector” and “consumer + behavior” queries, as demonstrated in Table 4.

Table 5 shows the enrichment of the query in terms of removing some concepts from the key term. Based on the KM in Fig. 9, COVID-19 has influenced the “logistics” part.
Table 3  Matrix of co-occurrence terms of key terms

| Key terms | Context terms  | $n$ | $\phi$ | Key terms | Context terms  | $n$ | $\phi$ |
|-----------|----------------|-----|--------|-----------|----------------|-----|--------|
| Food retail | Stores          | 1028 | 0.8740023 | Food service | Industry      | 1258 | 0.7256924 |
|           | Worker          | 985  | 0.8880104 | Government | 1012 | 0.8062149 |
|           | Grocery         | 973  | 0.8763271 | Consumers  | 856  | 0.6725691 |
|           | Foodstuffs      | 902  | 0.7659201 | Sector     | 823  | 0.7832588 |
|           | Online          | 753  | 0.8841263 | Workers    | 789  | 0.2362459 |
|           | Service         | 530  | 0.7506144 | Restaurant | 706  | 0.4936511 |
|           | Covid           | 423  | 0.3056318 | Products   | 526  | 0.5265426 |
|           | Sector          | 320  | 0.8740023 | Demand     | 302  | 0.3826559 |
|           | Employees       | 250  | 0.5756321 | Security   | 264  | 0.7269556 |
|           | Shopping        | 205  | 0.2036214 | Delivery   | 212  | 0.0936559 |
|           | E-commerce      | 105  | 0.1025964 | Waste      | 156  | $-0.2046952$ |
| Income    | 98              | 0.6584220 |          | Challenge  | 93   | $-0.1969524$ |
| Availability | 86             | $-0.1794235$ |      | Packaging  | 87   | 0.1795882 |
| Banks     | 52              | $-0.0856321$ |      | FAO        | 75   | $-0.0965289$ |
| Charge    | 25              | $-0.1794235$ |      | Consumption| 65   | $-0.1794235$ |
| Home      | 12              | $-0.0856321$ |      | Restaurant | 56   | $-0.0856321$ |
| Consumer  | Shopping        | 925  | 0.7856954 | Adapt      | 865  | 0.5629656 |
|           | Users           | 856  | 0.4852115 | Sector     | 725  | 0.2265995 |
|           | Law             | 802  | 0.69523485 | Product    | 702  | 0.0236955 |
|           | Spending        | 725  | 0.3958442 | Covid      | 602  | 0.1958521 |
|           | Behavior        | 652  | 0.7954215 | Covid      | 62   | 0.1958521 |
|           | Respondent      | 562  | 0.2953215 | Industry   | 406  | 0.2695225 |
|           | Enforcement     | 423  | 0.1695584 | Demand     | 369  | 0.1236955 |
|           | Food            | 365  | 0.2954225 | Automation | 302  | 0.6958848 |
| Live      | 302             | $-0.1558569$ |      | Growth     | 256  | 0.3695584 |
| Eating    | 258             | 0.5366549 |          | Impact     | 203  | 0.2998455 |
| Drinking  | 203             | 0.6569542 |          | Operation  | 123  | 0.7514523 |
| Crisis    | 195             | 0.1658447 |          | Companies  | 120  | 0.8541236 |
| Comes     | 102             | $-0.065948$ |      | Time       | 92   | $-0.8582263$ |
| Said      | 96              | $-0.3625529$ |      | Technology | 82   | 0.6592215 |
| Learning  | 85              | $-0.4595548$ |      | Serving    | 78   | $-0.5658412$ |
| Half      | 76              | $-0.2958525$ |      | Survive    | 65   | $-0.2325974$ |
| Logistic  | Market          | 852  | 0.2369584 |          |          |      |        |
|           | Costs           | 802  | 0.5265842 |          |          |      |        |
|           | Law             | 725  | 0.6589445 |          |          |      |        |
|           | Disruption      | 695  | 0.2569472 |          |          |      |        |
|           | Transport       | 510  | 0.3369525 |          |          |      |        |
|           | Management      | 453  | 0.3654258 |          |          |      |        |
|           | Processing      | 326  | 0.2568445 |          |          |      |        |
|           | Impact          | 237  | 0.6958444 |          |          |      |        |
| Survey    | 203             | $-0.5258447$ |      |          |          |      |        |
| Covid     | 126             | 0.2569577 |          |          |          |      |        |
| Firms     | 103             | 0.2569542 |          |          |          |      |        |
| Recovery  | 92              | 0.3256521 |          |          |          |      |        |
| Monitor   | 83              | $-0.9658255$ |      |          |          |      |        |
| Availability | 62             | $-0.3652458$ |      |          |          |      |        |
| Tool      | 52              | $-0.9655258$ |      |          |          |      |        |
One of the lower layer terms of this key term is “management.” The “logistics-management” query aims to examine the effects of COVID-19 on “logistics” without considering management concepts. Therefore, as shown in Table 5, the terms “truck,” “freight,” “impact,” “sea,” “consumer,” “transportation,” “frequency,” “congestion,” and “company,” which are more similar to this query and can enrich it, have been extracted. From the obtained terms, the terms “truck,” “freight,” “sea,” “consumer,” and “transportation” were selected for enrichment.

The terms vector with a higher analogy rate to the query is closer to the query vector than the other terms vector. In Fig. 10, the relationship between the two-dimensional representation of the query vector “food retail + online” and the term vector “channel; business; impact; COVID; consumer” were selected for enrichment.

The terms vector with a higher analogy rate to the query is closer to the query vector than the other terms vector. In Fig. 10, the relationship between the two-dimensional representation of the query vector “food retail + online” and the term vector “channel; business; impact; COVID; consumer” were selected for enrichment.

14 KUs. A set of KUs focusing on the enriched query revealed that the effects of the COVID-19 on the food service industry have been significant. A set of KUs showed that COVID-19 had greater influence over downstream retail and food service sectors, mostly informal SMEs.

Another constructed query was the “food retail + online.” Query enrichment operation was performed. The resulting query has been extracted as “food retail + online + channel + impact + consumer.” The graph of “Similarity value of extracted KU with the query” has five different clusters. The percentage of IV in each cluster is 52%, 25%, 16%, 6%, and 1%. Clusters 1 to 4 have high IV values and include 14 KUs. Some of the extracted KUs are shown in the output evaluated in the figure.

Another query was the “food retail + worker.” After enriching the query, the resulting query was extracted as “food retail + worker + company + impact + safety.” The graph of “similarity value of extracted KU with the query” has six different clusters. The percentage of IV of each cluster is 42%, 20%, 13%, 19%, 5%, and 1%, respectively. Clusters 1 to 5 have high IV, which include 14 KU. The extracted KUs showed that supermarket chains are already facing an increasing demand for workers and

Knowledge extraction/evaluation

Based on the created query, KUs were extracted from the corpus. Then, the information value of each KU was calculated, and the most valuable KUs were extracted. The resultant evaluated KUs are shown in Fig. 11, and the details of extracting them are described below:

- The query was selected as “food service + sector + restaurant.” After enriching the query, the resulting query was extracted as “food service + sector + restaurant + home + producer.”
- According to the percentage of IVs for each cluster, shown as a pie chart, with total values of 43%, 40%, 15%, and 2%, clusters 1 to 3 have more significant IVs with...
workers receiving low wages and inadequate social security benefits.

- Based on the KM, a query was created as “consumer + behavior,” focusing on the term “consumer” as a key term. According to the percentage of IVs of each cluster regarding the total value, 42%, 38%, 18%, and 2%, clusters 1 to 3 have significant IVs, including 12 KUs. The extracted KUs are shown in the evaluated output in the figure. According to the KUs extracted based on a query, it is clear that the rapid spread of COVID-19 and its impact has led to an increase in food health-conscious behavior. By selecting the query as “consumer + behavior + restaurant,” it was found that customers of restaurants and shopping malls have fallen sharply. Also, by creating a query called “consumer + behavior + buying,” the extracted KUs show that consumer buying behavior has become more erratic. Due to the pandemic crisis, the use of digital tools in customers’ buying patterns has increased and will probably continue after this crisis. Online food ordering has increased dramatically, and the desire to buy healthy products has increased among consumers. KUs, by focusing on “shop,” and selecting the query as “consumer + behavior + shop,” show that traditional consumers increasingly use the frictionless shopping method during the pandemic era. At the same time, several extracted KUs show that fast delivery of food by farm shops is an advantage, while during the pandemic, stores have not been able to cope with the high demand of customers.

- The query “logistic-management” was selected. To construct a suitable query, all terms similar to the “logistic-management” expression have been used to modify the query. The resulting query has been extracted as “freight + truck + sea + transportation.” According to the IV percentage of clusters, 1 to 3 have significant IVs, 31%, 29%, 28%, 10%, and 2%, respectively. Some of the extracted KUs are shown in the figure.

- Based on the search for KUs focusing on “logistic-management,” several KUs were extracted, implying that due to the increase in inspections of shipments, logistics costs have increased to comply with border control protocols during the pandemic. Also, based on the KUs extracted with the focus on “logistic-management + freight + truck + sea,” this hypothesis was obtained that, in the ocean freight sector, due to the emerged crisis, sailing programs are subject to disruption, and ports and terminals are facing equipment imbalances. Operational restrictions have also led to delays in the delivery of goods, congestion, and rising final prices.

- Another selected query was the “manufacture + operation.” The enriched query “manufacture + operation + plants + products + demand + packaging” was constructed based on this query. The evaluated KUs.
show that clusters 1 to 3 have significant IVs, 45%, 32%, 21%, and 2%, respectively. The extracted KUs based on this query showed that sales of plant-based products had increased drastically during the pandemic. In the KUs discovered with the terms “demand + packaging” in the lower layers of the KM, the obtained KUs showed that food suppliers faced a sudden increase in sales, during the pandemic, due to the increased demand of consumers for packaged and processed foods. The demand for food robotics has also increased, emphasizing social distancing considerations.

**Discussion**

The main goal of this study was to apply the KM and NLP techniques to identify the effects of natural crises on the SC. To achieve this goal, we proposed and implemented a framework. After applying topic modeling and analyzing word frequency and co-occurrences on 975 analytical reports related to the effects of the natural crisis on the SC, a KM of the five main related topics was created and expanded. This study used the KM (or co-occurrence network) of risk terms to derive and categorize 53 knowledge propositions, which express the impact of the COVID-19 crisis on essential parts of the SC in a categorized manner.

The extracted categorized KUs from user-system interactions could be used to inform industry experts, governments, and academic researchers about the various effects of natural crises on the SC. Natural crises are disruptive and affect SC performance. Therefore, effective SC crisis monitoring is necessary to reduce costs and increase the organization’s sustainability in the long term. The proposed framework can help organizations automatically develop and expand a hierarchical KM of crisis effects. In addition, they can target critical issues judged by analysts to monitor risks, identify threats, and determine a corrective or protective course of action. The extracted terms of SC crises make a dictionary of crises that can be used to create a suitable query to extract knowledge propositions from analytical reports related to various SC crises.

The resultant KM and KUs show five sectors of the pandemic’s impact: (1) food retail, (2) food services, (3) manufacture, (4) consumer, and (5) logistics, briefly described as follows:

- In the consumer sector: The consumption patterns have been changed during the COVID-19 pandemic.
Fig. 11 Evaluated knowledge units outputs

- In the manufacturing sector: The consumer demands have been increased for packaged and processed foods; robots have been more used in the production process; and healthier materials such as vegetables, fresh, and healthier foods have been more produced.
- In the logistic sector: Due to the increase in inspections of shipments, logistics costs have increased to comply with border control protocols during the pandemic.
- In the food retail and the food services sectors: Various aspects have been affected by the pandemic crisis, such as increased demand for workers in supermarkets, drastic changes in consumption patterns, and food ordering. Providing such categorized knowledge propositions as a decision support system can assist SC analysts in using such a KM in short and long-term decisions in similar natural crises.

By using text-mining capability in categorizing large volumes of data, this research solves the problem of validation of qualitative methods in crisis factors analysis. On the other hand, using human intelligence in steering text-mining algorithms to correct the obtained results has solved the problems in text analysis methods to analyze complex crisis data in related research.

Conclusion

The main goal of this research is to propose a text-analysis-based framework using co-occurrence term analysis and knowledge map construction to analyze a large volume of textual data, retrieved from websites and social media, concerning crisis effects on supply chains (SCs).
We implemented the proposed framework, and to validate the framework, we conducted a case study of the effects of COVID-19 on SCs. For this purpose, 975 online analytical reports on related topics were gathered and cleaned as a corpus. Then, data pre-processing operations were performed to prepare the data for further processing. And finally, using topic modeling, term frequency, and co-occurrence analysis, a set of analytical reports related to the effects of natural crises on the SC was analyzed. By textual analysis, 53 crisis knowledge propositions were extracted from 25,272 sentences in the corpus. Also, a knowledge map was created by co-occurrence analysis from 631,799 terms and 31,864 unique terms that identify the relationship between the 110 main crisis factors.

Using the implemented framework, a KM of crisis effects and a dictionary of crisis factors on the SC were created in five sectors: (1) food retail, (2) food services, (3) manufacture, (4) consumer, and (5) logistics. This KM made it possible to create a suitable query for extracting categorized SC crisis knowledge. This categorization can improve the decision-makers’ knowledge in dealing with SC crises.

According to an expert panel, the categorized KUs have been monitored and reported the effects of the COVID-19 crisis on the SC in various sectors. In this research, using the implemented framework, we created a crisis KM and related knowledge extraction through user steering to identify the effects of the natural crisis on the SC.

**Theoretical implication**

This paper addresses two main research gaps. First, much research has been done to identify SC crises based on qualitative methods. These methods, in addition to sample limitations, are subject to human biases and judgments that cause results to be questionable. For example, survey data in research by Giannakis et al. (2016), based on human judgments, raise concerns about possible biases. Second, data-driven research has been conducted by analyzing large volumes of data to minimize the problem of data limitations and human judgments. These methods analyze the data using conventional text analysis methods, regardless of the nature of the risk (Wang and Ye 2018; Kara et al. 2020; Ganesh and Kalpana 2022). Risk data are ambiguous, complex, and have multiple meanings; thus, one risk term may fall into different sectors (Pai et al. 2003; Chiu and Choi 2016). Mere text analysis methods cannot make this distinction and require human steering (Jin et al. 2018; Chu et al. 2020).

Our proposed method could be helpful to address both mentioned gaps. Since our method applies text-mining algorithms to analyze a large volume of SC crisis data, the sample limitation could be addressed. Moreover, our proposed method is based on user-system interactions; therefore, the complexity and vagueness of data could be resolved due to user feedback. To the best of our knowledge, no previous research has used this method for risk analysis, including analyzing the effects of natural crises on the SC and to be used in decision support systems.

Thus, very briefly, this study significantly extends the existing literature on understanding the effects of the crisis on SC in two aspects: (1) using large volume of data and therefore resolving the sample limitation in the qualitative method and (2) enhancing the analysis performance, benefiting from the user steering in the framework.

**Practice implications**

SC crisis monitoring framework could be useful for both firms and governments. The firms could use this framework for their business continuity, and governments could use it for their effective policymaking (O’Rourke 2014). The current literature shows how specific decision-making processes before, during, and after a crisis can improve a company’s SC performance (Dayton and Bernhardsdottir 2010). The development of information networks has enabled companies to access and analyze crisis information transparently. Companies can make decisions and take action using crisis data analysis instead of evidence-based speculation (Jin et al. 2018; McAfee et al. 2012). Data analysis is used for various operations such as procurement, service, production, warehousing, and demand management in SC management. For instance, Nguyen et al. discussed using metadata in SC management (Nguyen et al. 2015).

Considering the importance of SC crisis management, this study proposes a framework for monitoring and analyzing crisis data and establishing a crisis decision support system for firms, enabling them to respond appropriately to crises that disrupt SC performance. This framework combines human skills with text analysis techniques to provide efficient crisis data management, categorized related knowledge, and evidence-based decision-making.

**Policy implication**

Applying this framework can also be very useful for government policymakers to provide effective policies to support consumer services and maintain the continuity of the business market. Natural crises are often characterized by rapid expansion and affect different parts of the SC (Natarajarathinam et al. 2009); therefore, governments pursue prudent policies to limit its spread (Barnes and Oloruntoba 2005). Given the impact of these policies on the performance of all parts of the SC, creating a categorized knowledge database about the SC crisis can enhance policymakers’ knowledge to make more effective policies (Miroudot 2020). The large
amount of data generated in information networks about SC crises has created a good opportunity for government policymakers to make evidence-based policies instead of intuition (Newig et al. 2016). Therefore, one of the objectives of this study is to propose a new decision support system that can guide policymakers through the analysis of crisis data and the creation of classified knowledge to create policies for SC crisis management. The extracted pieces of knowledge are used to inform advice, monitor, evaluate, and revise the decisions made by policymakers.

Limitations and future research

There are several limitations in the study which could be addressed in future research. First, the KM of crisis does not discover the type of relationship between crisis terms. Crisis terms are related to each other, and identifying these relationships can be very useful in constructing the appropriate queries. Second, the subject was limited to social media reports for analysis in the data collection phase. This is just for the purpose of representing the operation of the proposed framework. Various automation systems such as enterprise resource planning and customer relationship management in companies collect very important applicable data related to financials, SC, operations, commerce, reporting, manufacturing, and human resource activities that should be considered in the analysis.

Hence, future developments of this research could focus on two aspects. First, it could focus on the analysis of dependencies among various crisis factors. Understanding the relationship between different types of crises significantly improves supply chain crisis classification and management. This can be achieved by creating ontology between the crisis factors. Secondly, for crisis analysis of various crisis data types such as automation systems, the future extension could focus on creating data analysis rules, identifying thresholds, and mapping the results to identify crisis factors in the supply chain.

Author contribution

Conceptualization: Mohammad Reza Sheikhattar, Navid Nezafati; Sajjad Shokouhyar; methodology: Mohammad Reza Sheikhattar, Navid Nezafati; formal analysis: Mohammad Reza Sheikhattar; investigation: Mohammad Reza Sheikhattar, Navid Nezafati, Sajjad Shokouhyar; data curation: Mohammad Reza Sheikhattar, Navid Nezafati; software: Mohammad Reza Sheikhattar, Navid Nezafati; visualization: Navid Nezafati; validation: Mohammad Reza Sheikhattar; writing original draft: Mohammad Reza Sheikhattar; reviewing: Mohammad Reza Sheikhattar.

Data availability

Not applicable.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent to publish Not applicable.

Competing interests The authors declare no competing interests.

References

Aday S, Aday MS (2020) Impact of COVID-19 on the food supply chain. Food Q Saf 4:167–180
Aghion P, Burgess R, Redding SJ, Zilibotti F (2008) The unequal effects of liberalization: evidence from dismantling the License Raj in India. Am Econ Rev 98:1397–1412
Ahmadi M, Sharifi A, Khalili S (2021) Presentation of a developed sub-epidemic model for estimation of the COVID-19 pandemic and assessment of travel-related risks in Iran. Environ Sci Pollut Res 28:14521–14529
Akundi A, Tseng B, Wu J, Smith E, Subbalakshmi M, Aguirre F (2018) Text mining to understand the influence of social media applications on smartphone supply chain. Procedia Comput Sci 140:87–94
Arlinghaus JC, Zimmermann M, Zahner M (2020) The influence of cognitive biases on supply chain risk management in the context of digitalization projects. In: Freitag M, Haasis HD, Kotzab H, Pannek J (eds) Dynamics in Logistics. LDIC 2020. Lecture Notes in Logistics. Springer, Cham. https://doi.org/10.1007/978-3-030-44783-0_13
Bakhitn P, Khabirova E, Kuzminov I, Thurner T (2020) The future of food production—a text-mining approach. Technol Anal Strat Manag 32:516–528
Barnes P, Oloruntoba R (2005) Assurance of security in maritime supply chains: conceptual issues of vulnerability and crisis management. J Int Manag 11:519–540
Baryannis G, Validi S, Dani S, Antoniou G (2019) Supply chain risk management and artificial intelligence: state of the art and future research directions. Int J Prod Res 57:2179–2202
Bevilacqua M, Ciaparica FE, Marcucci G, Mazzuto G (2020) Fuzzy cognitive maps approach for analysing the domino effect of factors affecting supply chain resilience: a fashion industry case study. Int J Prod Res 58:6370–6398
Blei DM (2012) Probabilistic topic models. Commun ACM 55:77–84
Braun V, Clarke V (2006) Using thematic analysis in psychology. Qual Res Psychol 3:77–101
Cao L, Li F-F (2007) Spatially coherent latent topic model for concurrent segmentation and classification of objects and scenes. IEEE 11th International Conference on Computer Vision, pp 1–8. https://doi.org/10.1109/ICCV.2007.4408965
Cerruti C, Delbufalo E (2009) International sourcing effectiveness in the fashion industry: the experience of Italian industrial districts. Int J Glob Small Bus 3:427–440
Chiu C-H, Choi T-M (2016) Supply chain risk analysis with mean-variance models: a technical review. Ann Oper Res 240:489–507
Chiu M-C, Lin K-Z (2018) Utilizing text mining and Kansei Engineering to support data-driven design automation at conceptual design stage. Adv Eng Inform 38:826–839
Chu C-Y, Park K, Kremer GE (2019) Applying text-mining techniques to global supply chain region selection: considering regional differences. Procedia Manuf 39:1691–1698

Chu C-Y, Park K, Kremer GE (2020) A global supply chain risk management framework: an application of text-mining to identify region-specific supply chain risks. Adv Eng Inform 45:101053

Da Silva JBN, Senna P, Chousa A, Coelho O (2020) Data mining and operations research techniques in supply chain risk management: a bibliometric study. Braz J Oper Prod Manag 17:1–14

Dang S, Ahmad PH (2014) Text mining: techniques and its application. Int J Eng Technol Innov 1:22–25

Dayton BW, Bernhardsdottir A (2010) Crisis management. In: Yong Dang S, Ahmad PH (2014) Text mining: techniques and its application. Elsevier 825–832. https://doi.org/10.1016/B978-0-08-092819-9.00084-6

Dayton BW, Bernhardsdottir A (2010) Crisis management. In: Yong Dang S, Ahmad PH (2014) Text mining: techniques and its application. Elsevier 825–832. https://doi.org/10.1016/B978-0-08-092819-9.00084-6

Dowling M, Wycoff N, Mayer B, Wenskovitch J, House L, Polya N, North C, Hauck P (2019) Interactive visual analytics for sensemaking with big text. Big Data Res 16:49–58

Elbattah M, Arnaud E, Gignon M, Dequen G (2021) The role of text analytics in healthcare: a review of recent developments and applications. 825–832. https://doi.org/10.5220/00010414508250832

Freilich S, Kreimer A, Meilijson I, Gophna U, Sharan R, Ruppin E (2010) The large-scale organization of the bacterial network of ecological co-occurrence interactions. Nucleic Acids Res 38:3857–3868

Ganesh AD, Kalpana P (2022) Supply chain risk identification: a real-time data-mining approach. Ind Manag Data Syst 122(5):1333–1354. https://doi.org/10.1108/IMDS-11-2021-0719

Giannakis M, Puest G, MacQueen KM, Namey EE (2012) Applied thematic analysis. SAGE Publications, Inc. https://doi.org/10.4135/9781843384436

Gudivada VN, Rao DL, Gudivada AR (2018) Information retrieval: concepts, models, and systems. In: Gudivada VN, Rao C (ed) Handbook of Statistics, Elsevier 38:331–401. https://doi.org/10.1016/B978-0-444-63410-7.00070-9

Guérin PJ, Singh-Phulengda S, Strub-Wourgaft N (2020) The consequence of COVID-19 on the global supply of medical products: Why Indian generics matter for the world? 9:225. https://doi.org/10.1002/11109.20200110263

Guest G, MacQueen KM, Namey EE (2012) Applied thematic analysis. SAGE Publications, Inc. https://doi.org/10.4135/9781843384436

Hu F, Zhang Y-F (2010) Text mining based on domain ontology. International Conference on E-Business and E-Government, pp 1456–1459. https://doi.org/10.1109/ICEE.2010.370

Jin M, Wang Y, Zeng Y (2018) Application of data mining technology in financial risk analysis. Wireless Pers Commun 102:3699–3713

Jing L-P, Huang H-K, Shi H-B (2002) Improved feature selection approach TFIDF in text mining. Proceedings. International Conference on Machine Learning and Cybernetics 2:944–946. https://doi.org/10.1109/ICMLC.2002.1174522

Jurafsky D, Martin J (2008) Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition, and Speech Recognition. Prentice Hall

Kam BH, Chen L, Wilding R (2011) Managing production outsourcing risks in China’s apparel industry: a case study of two apparel retailers. Supply Chain Management 16(6):428–445. https://doi.org/10.1108/13598541111171147

Kara ME, Firat SüO, Ghadge A (2020) A data mining-based framework for supply chain risk management. Comput Ind Eng 139:105570

Karwasra K, Soni G, Mangla S, Kazancoğlu Y (2021) Assessing dairy supply chain vulnerability during the Covid-19 pandemic. Int J Logist Res Appl 1–19. https://doi.org/10.1080/13675567.2021.1910221

Köksal D, Strähle J, Müller M (2018) Social sustainability in apparel supply chains—the role of the sourcing intermediary in a developing country. Sustainability 10:1039

Maillot DGC, Wiber MG, Barnett A (2019) Actions towards the joint production of knowledge: the risk of salmon aquaculture on American Lobster. J Risk Res 22:67–80

Martino G, Fera M, Iannone R, Miranda S (2017) Supply chain risk assessment in the fashion retail industry: an analytic network process approach. Int J Appl Eng Res 12:140–154

Mcafee A, Brynjolfsson E, Davenport TH, Patil D, Barton D (2012) Big data: the management revolution. Harv Bus Rev 90:60–68

Menut L, Bessagnot B, Siour G, Mailler S, Pennel R, Cholakian A (2020) Impact of lockdown measures to combat COVID-19 on air quality over western Europe. Sci Total Environ 741:140426

Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. CoRR, abs/1301.3781. https://arXiv.org/10.48550/arXiv.1301.3781

Mirodout S (2020) Reshaping the policy debate on the implications of COVID-19 for global supply chains. J Int Bus Policy 3:430–442

Momani A, Rost K (2016) Identification and monitoring of possible disruptive technologies by patent-development paths and topic modeling. Technol Forecast Soc Chang 104:16–29

Moon K, Chan RL, Davis BL (2010) Adoption of enterprise risk management: A study of a textile and clothing supply chain. 8th International Conference on Supply Chain Management and Information, pp 1–4

Nagamachi M (1995) Kansei engineering: a new ergonomic consumer-oriented technology for product development. Int J Ind Ergon 15:3–11

Nasirpour MH, Sharifi A, Ahmadi M, Jafarzadeh Ghoushchi S (2021) Revealing the relationship between solar activity and COVID-19 and forecasting of possible future viruses using multi-step autoregression (MSAR). Environ Sci Pollut Res 28:38074–38084

Natarajathanam M, Capar I, Narayanan A (2009) Managing supply chains in times of crisis: a review of literature and insights. Int J Distrib Logist Manag 39(7):535–573. https://doi.org/10.1108/09600030910996251

Nazam M, Hashim M, Randhawa M, Maqbool A (2020) Modeling the barriers of sustainable supply chain practices: a Pakistani perspective. https://doi.org/10.1007/978-3-030-21255-1_27

Newig J, Kochskämper E, Challies E, Jager NW (2016) Exploring governance learning: how policymakers draw on evidence, experience and intuition in designing participatory flood risk planning. Environ Sci Policy 55:353–360

Nguyen PH, Xu K, Wheat A, Wong BW, Attfield S, Fields B (2015) Sensepath: Understanding the sensemaking process through analytical provenance. IEEE Trans Visual Comput Graphics 22:41–50

O’rourke D (2014) The science of sustainable supply chains. Science 344:1124–1127

Pai RR, Kallepalli VR, Caudill RJ, Zhou M (2003) Methods toward supply chain risk analysis. SMC’03 Conference Proceedings. IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance (Cat. No.03CH37483) 5:4560–4565. https://doi.org/10.1109/ICSMC.2003.1245702

Palmer D, Mavroidis P,Meagher N (2022) Dispute settlement in the World Trade Organization: Practice and Procedure. In: Dispute Settlement in the World Trade Organization (p. II). Cambridge: Cambridge University Press

Ponis ST, Ntalla A (2016) Crisis management practices and approaches: insights from major supply chain crises. Procedia Econ Finance 39:668–673

Pulijc M (2010) The influence of cognitive biases on managerial perceptions of supply chain risk. In: Dani S (ed) Proceedings of the the 10th International Research Seminar on Supply Chain Risk Management. Loughborough University, UK

 Springer
Qaiser S, Ali R (2018) Text mining: use of TF-IDF to examine the relevance of words to documents. Int J Comput Appl 181:25–29
Rizou M, Galanakis IM, Aldawoud TM, Galanakis CM (2020) Safety of foods, food supply chain and environment within the COVID-19 pandemic. Trends Food Sci Technol 102:293–299
Schorsch T, Wallenburg CM, Wieland A (2017) The human factor in SCM: Introducing a meta-theory of behavioral supply chain management. Int J Phys Distrib Logist Manag 47(4):238–262. https://doi.org/10.1108/IJPDLM-10-2015-0268
Segev E (2020) Textual network analysis: detecting prevailing themes and biases in international news and social media. Sociol Compass 14:e12779
Segev E (2021) Semantic network analysis in social sciences. Routledge, London
Shah SM, Lütjen M, Freitag M (2021) Text mining for supply chain risk management in the apparel industry. Appl Sci 11:2323
Shanker S, Barve A, Muduli K, Kumar A, Garza-Reyes J A, Joshi S (2021) Enhancing resiliency of perishable product supply chains in the context of the COVID-19 outbreak. Int J Logist Res Appl. https://doi.org/10.1080/13675567.2021.1893671
Sharifi A, Ahmadi M, Ala A (2021) The impact of artificial intelligence and digital style on industry and energy post-COVID-19 pandemic. Environ Sci Pollut Res 28:46964–46984
Sharma M, Luthra S, Joshi S, Kumar A (2020) Developing a framework for enhancing survivability of sustainable supply chains during and post COVID-19 pandemic. Int J Logist Res Appl. https://doi.org/10.1080/13675567.2020.1810213
Spitsberg I, Brahmandam S, Verti MJ, Coustun GW (2013) Technology landscape mapping: at the heart of open innovation. Res Technol Manag 56:27–35
Su C-J, Chen Y-A (2018) Risk assessment for global supplier selection using text mining. Comput Electr Eng 68:140–155
Trautrim A, Schleper MC, Cakir MS, Gold S (2020) Survival at the expense of the weakest? Managing modern slavery risks in supply chains during COVID-19. J Risk Res 23:1067–1072
Wan Q, Xu X, Zhuang J, Pan B (2021) A sentiment analysis-based expert weight determination method for large-scale group decision-making driven by social media data. Expert Syst Appl 185:115629
Wang Z, Ye X (2018) Social media analytics for natural disaster management. Int J Geogr Inf Sci 32:49–72
Yan W, He J, Trappey A (2019) Risk-aware supply chain intelligence: AI-enabled supply chain and logistics management considering risk mitigation. Adv Eng Inform 42:100976. https://doi.org/10.1016/j.aei.2019.100976

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.