



Blockchain technology in supply chain management: insights from machine learning algorithms

Hirata, Enna
Lambrou, Maria
Watanabe, Daisuke

(Citation)

Maritime Business Review, 6(2):114-128

(Issue Date)

2021-05-21

(Resource Type)

journal article

(Version)

Accepted Manuscript

(Rights)

This article is © 2020, PACIFIC STAR GROUP EDUCATION FOUNDATION and permission has been granted for this version to appear here <http://www.lib.kobe-u.ac.jp/kernel/>. Pacific Star Group Education Foundation does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission of...

(URL)

<https://hdl.handle.net/20.500.14094/90008813>





Applications of Blockchain Technology in Supply Chain Management: Insights from Machine Learning Algorithms

Journal:	<i>Maritime Business Review</i>
Manuscript ID	MABR-07-2020-0043.R3
Manuscript Type:	Research Paper
Keywords:	Blockchain, Supply Chain, Machine Learning, word2vec, LDA, Natural Language Processing (NLP)

SCHOLARONE™
Manuscripts

Applications of Blockchain Technology in Supply Chain Management: Insights from Machine Learning Algorithms

1. Introduction

The application of blockchain in the supply chain sector has been vastly discussed recently from both optimistic and negative perspectives. Many consider applying blockchain in supply chain and logistics is an exciting prospect, there are also concerns that the technology lacks the maturity (Duru and Zin, 2019) today to handle global supply chain complexity. Such limitations might be discouraging, but as an innovative technology, blockchain has the potential to reform global supply chains. Practitioners should meticulously consider its existing hurdles and possible challenges as the technology matures. A number of multi-case, qualitative research works proposed the overarching theoretical model, which systematizes the technological components, the prevailing management rationales and determinant factors of digitalization including blockchain (Lambrou et al., 2019a, 2019b; Wagner and Wiśnicki, 2019; Duru and Zin, 2019). However, these case studies are basically based on deductive method and it is difficult to ensure generality.

Academic literature has become a rich source of information for researchers, practitioners and informed citizens, on various technological applications. Researchers use documents to express new ideas, theories, hypotheses, methods, approaches, and experimental results with other researchers and interested parties. Text documents, such as research articles, technical reports, and patents, are the preferred method of communication by researchers. Therefore, there is a lot of effort put into scientific communication, with scientific texts presenting a challenge to text mining methods since the language used is formal and highly specialized.

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 the processing of extracting information from a large set of documents more efficient. However, since natural language is not easily processed by computer programs, it is necessary to develop algorithms to transform text into a structured representation, which is performed through Natural Language Processing (NLP).

NLP is a branch of artificial intelligence that helps computers to understand, interpret and manipulate human natural languages. For example, “NLP makes it possible for computers to read text, hear speech, interpret language, 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, 2020).

Increasingly, nowadays, experimental, emergent and mixed methods research approaches are acknowledged as constituting valid academic discourse choices, in their different nuances and sophistication (Eickhoff and Neuss, 2017; Asmussen and Charles, 2019). Both blockchain

researchers and supply chain management scholars have started to embrace topic modeling as a promising research method.

In our paper, we employ a complementary, alternative perspective to investigate blockchain application in supply chains, based on text mining techniques for academic literature content analysis. The mined text is then applied with a couple of machine learning models to extract useful information.

Previous research papers have mainly analyzed bibliometric or abstract of scientific articles. Different from the approaches existing literature took, this study presents a method of applying NLP to extract information from the full text of scientific articles. The extracted text documents are then trained by a machine learning algorithm which performs automatic text classification. The rest of the paper is organized as follows: section 2 reviews literature; section 3 outlines the methodology; section 4 discusses the results and findings; section 5 concludes the research and future prospects.

2. Literature Review

The multi-faceted phenomenon of blockchain technology pertaining supply chains has recently attracted considerable academic research attention and efforts.

The first stream of academic papers focuses on exemplifying both technology features and business incentives to adopt blockchain in supply chains, along with enabling and constraining factors. Sternberg and Baruffaldi (2018) reviewed supply chain blockchain initiatives and theorized the logic and challenges of blockchains in the supply chain industry. The authors concluded that while several incentives of developing and using blockchain technology exist, it was not apparent how companies can actually benefit with a materialized business advantage.

Kshetri (2018) also reviewed early supply chain industry cases and delineated the blockchain's role in supply chain management. The cases illustrate drivers and mechanisms for meeting cost, quality, speed, dependability, risk reduction, sustainability and flexibility objectives of supply chain organizations. Furthermore, identified determinants of blockchain adoption include the number of entities involved (viable blockchain ecosystem), participants' capabilities and the extent of industry competitive pressure.

Saberi et al. (2019) overviewed blockchain technology and its applicability in the supply chain, systematized a comprehensive list of barriers (i.e. system-related, intra and inter-organizational and external barriers) and proposed certain directions for overcoming those salient obstacles (i.e. governance mechanisms).

The second stream of research works endeavors to analyze how the current blockchain technology applications in supply chain management are implemented, delineate the foundations of the technology and further articulate the value of the technology for supply chain management, towards identifying pertinent enablers for achieving certain business goals and broader market adoption (Blossey et al., 2019; Casey and Wong, 2017; Queiroz et al., 2019; Roeck et al., 2020; Wang et al., 2019). Gurtu and Johny (2019) also discuss the significance and

applications of blockchain technology with elaborate references to both generic supply chain, transport and maritime logistics cases.

A number of papers examine specific supply chain applications, such as food or pharmaceutical supply chain blockchains, shedding light on particular aspects, such as determinants to achieve visibility and trust via blockchain (Rogerson and Parry, 2020), or the interplay of blockchain technology features (i.e. consensus protocols) and business model requirements and blockchain ecosystem governance.

Research efforts focusing on specific supply chain areas, such as ports and shipping are emerging as well. Tsiulin et al. (2020) categorize blockchain projects in shipping and supply chain management and discuss the interrelations between blockchain features and shipping and ports concepts, towards delivering a better understanding of suitable use scenarios.

The third stream of literature re-employs present management theory in order to comprehend the unfolding of supply chain transformation resultant from blockchain technology; Treiblmaier (2018) employs core economics and management theories, namely principal agent theory, transaction cost analysis resource-based view and network theory, in order to address the implications stemming from applying blockchain in supply chain management. Roeck et al. (2020) examine in particular how blockchain technology affects transactions and governance modes in supply chains, viewed from a transactions cost economics point of view, while conducting an abductive multiple case study of five supply chain industry cases. Kummer et al. (2020) identify pertinent organizational theories used in blockchain literature in the context of supply chain management, in specific agency theory, information theory, institutional theory, network theory, the resource-based view and transaction cost analysis. Most importantly, the authors reframe supply chain management research questions addressed from the identified organizational theories vantage point, as intertwined with blockchain technology.

The latest, more mature stream of academic research, now habitually, focusses on more fine-grained topics, such as feasibility assessment and decision making for technology selection (Ar et al., 2020). Bai and Sarkis (2020) assess the growing literature studying the application of blockchain technologies in supply chain management; their research findings also identify a broad range of application types and operational objectives pursued (i.e. traceability, avoiding counterfeit products or reducing carbon footprints) and associated blockchain technical characteristics (i.e. scalability, complexity, security etc.). Furthermore, the authors propose a performance measures framework that considers how blockchain technologies can help supply chain meet targeted key objectives, based on hesitant fuzzy set and regret theory.

Apparently, maritime and shipping research is indeed rejuvenated in terms of both methods and research topics. Fiskin and Cerit (2020), apply bibliometric and network analysis towards identifying areas of current research interests in the entire body of shipping literature, revealing interesting publication clusters, their relationships and changes over five years. Lee and Shin (2019) apply topic modeling on port research publications. Shin et al. (2018) conducted a literature review study on sustainability in maritime research, with text mining.

Chang and Chen (2020) provide an elaborate review of recent blockchain studies in the supply chain management context, systematizing an enhanced list of topics and applications.

The identification of research problems, appropriate theories and methods to investigate the application of blockchain technology in supply chain management is predominantly conducted according to the social sciences, positivist, empirical research tradition. Nonetheless, alternative research approaches are also used, beyond qualitative or directed content analyses and case study field research designs. Pournader et al. (2020) review the existing academic literature and industrial knowledge sources regarding the applications of blockchains in supply chains, logistics and transportation, and identify the 4Ts – technology, trust, trade, traceability/transparency (research themes clusters), based on a co-citation analysis of the publications on this topic. Wang et al. (2018) compare and summarize 29 of blockchain studies in logistics area, and suggest value of blockchain in four areas, namely extended visibility and traceability, supply chain digitalization and disintermediation.

Against this background, to the best of our knowledge, extant literature examining blockchain application to supply chain management has indeed reached a maturity level where a sufficient number of pertinent research questions (or themes), revealed by multidisciplinary research frameworks examining blockchain technology and its implications, have been brought into the supply chain management academic discourse with varying intensity, rigor and insight (Iansiti, and Lakhani, 2017).

Currently, we do have a fair understanding of how, in particular, supply chain blockchain design and features (i.e. consensus mechanisms, security configurations, immutability, decentralized control) impact:

- (i) organizations and industries, in specific how blockchain enabled new business models are unfolding in the supply chain sector, and how different intertwined industries comprise the disruptive potential that blockchain technology involves
- (ii) platforms, different blockchain implementations and protocols (i.e. Hyperledger), as well as various types of supply chain blockchains (i.e. private, permissioned), as well as inter-platform interoperability and integration with legacy systems are unfolding
- (iii) intermediation, the manner alternative blockchain features and designs enact different intermediation possibilities i.e. complementing existing supply chain intermediaries rather than excluding them
- (iv) users and society, in particular how growing market adoption is shaping societal effects, such as sustainable development goals, and eventually realize and enact a multiplicity of possibilities regarding how supply chain blockchains create value (Risius and Spohrer, 2017).

Asmussen and Charles (2020) identified the current state-of-the-art digital technologies in supply chain management, and enablers for competitive advantage, based on a topic modelling framework. Shahid (2020) derived Latent Dirichlet Allocation (LDA) topics of blockchain

research (i.e. novelty, disruption, business blockchain types, protocol development etc.), contributing with an efficient reporting of research trends and identified potential areas for interdisciplinary blockchain research collaboration. LDA has been applied in maritime related studies which generated useful insights. Shin et al. (2018), Lee and Shin (2019) apply LDA to identify research topics and suggest future research should focus on port collaboration and environmental issues.

The existing literature has mainly analyzed bibliometric data and/or abstracts, few of them have analyzed the paper contents. By analyzing full text articles, our paper aspires to contribute in the advancement of supply chain management and maritime transport research with a particular application of machine learning techniques to blockchain literature review and rigorous, new theory building. We exemplify our model as situated in the interface of human-based and machine-learned analysis, potentially offering an interesting and relevant avenue for blockchain and supply chain management researchers.

3. Methodology

The proposed method is organized into three major modules namely, pre-processing, machine learning, and visualization. The pre-processing stage involves the techniques and processes which conducts the task of text mining. A couple of machine learning models (PCA, word2vec, LDA) are formulated by the training modules, which conduct the learning and classification tasks. Finally, the visualization phase describes the findings of the study. The workflow of the proposed system is represented as follows.

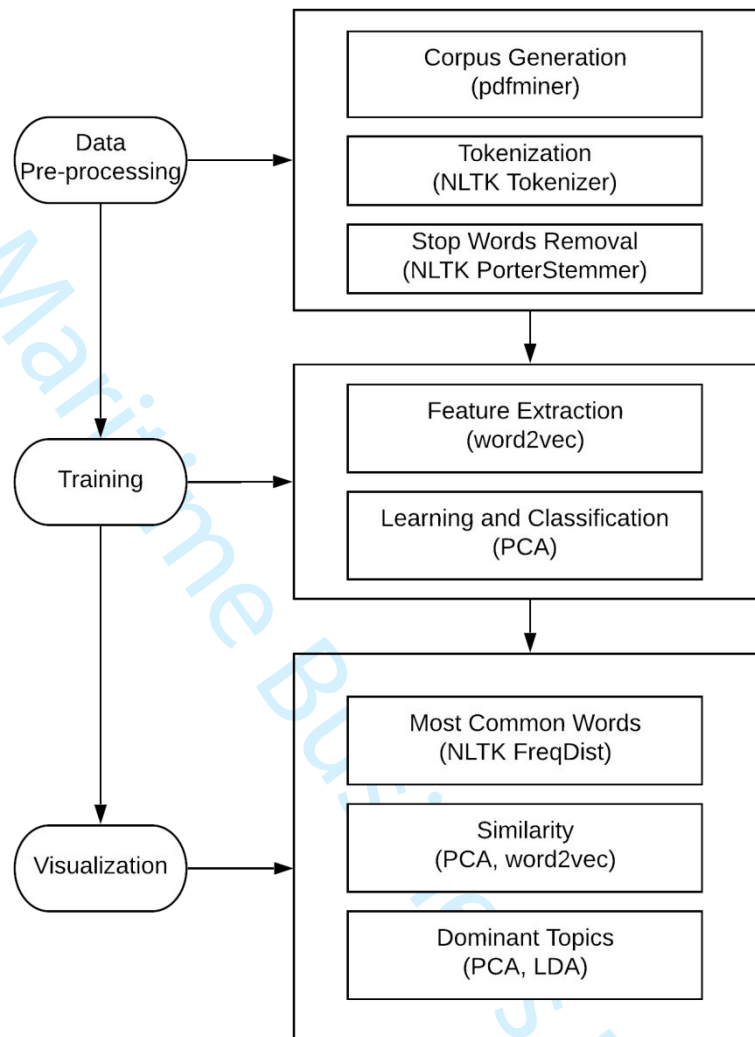


Figure 1 – Proposed Method (tools utilized are indicated in brackets)

3.1 Data Pre-processing

3.1.1 Text mining

Text analysis allows automatic extraction and classification of information from text (e.g. Westergaard et al, 2018), such as tweets, emails, product reviews, and survey responses. Popular text analysis techniques include word frequency, collocation, concordance, text classification, sentiment analysis, topic detection, language detection, clustering, keyword extraction and entity recognition etc.

Sorting through data is a repetitive, time-consuming and expensive process if done by humans. Instead, if done by machines, high volumes of text can be analyzed with least efforts, while providing even more accurate insights.

Text mining in this study is performed in following steps.

3.1.2 Corpus Generation

The experiment of this study is carried out on a text corpus which is a collection of literature published in Science Direct, Emerald and Springer database with following criteria. Time span is set to be from January 1990 to January 2020. The articles are retrieved by using key words “blockchain” and “supply chain”. In total, 422 articles hit the search. Table 1 outlines the distribution of corpora. Only literature in English language is included in this study.

Table 1 – Corpora

Corpus	Numbers of Related Article	Document Types
Science Direct	68	Full text articles
Emerald	103	Full text articles
Springer	251	Full text articles
Total	422	Full text articles

Since the retrieve articles are in pdf format, we use python script pdfminer to convert the pdf data to text format for further analysis. There are other similar tools available, which should generate same result as what pdfminer does in converting pdf data to text format.

3.1.3 Tokenization

Tokenization is a critical and the most basic step to proceed with NLP. Tokenization in NLP means to split raw text into smaller units, such as words or terms, which are called tokens. These tokens are the key elements of the NLP.

Tokenization plays an important role in NLP because tokenization provides a way to easily interpret the meaning of a text by analyzing the sequence of the words in the text. To have a better understanding of tokenization, let’s consider the below sentence:

“Blockchain and supply chain are a match made in heaven.”

Tokenize the sentence, we will get:

[‘blockchain’, ‘and’, ‘supply’, ‘chain’, ‘are’, ‘a’, ‘match’, ‘made’, ‘in’, ‘heaven’]

The python script of NLTK (Natural Language Tool Kit) Tokenizer is applied in this study to split the text data to tokens. The corpus applied in this study contains 4,775,532 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.

3.1.4 Stop Words Removal

Next stage of data pre-processing is stop words removal. Stop words are words which are commonly used in any natural language. For the purpose of analyzing text data and building NLP models, those stop words might not add much value to the meaning of the document, as such they are often filtered out in the data pre-processing stage.

Stop words usually refers to the most common words in a language, however, there is no single universal list of stop words. In this study, two types of stop words are removed: common English stop words (e.g. 'is', 'was', 'where', 'the', 'a', 'for', 'of', 'in') and some extra stop words (e.g. 'ieee', 'paper', 'vol', 'doi', 'et', 'al', 'https', 'www' etc.) that are associated with the corpus particularly. Removing these stop words help reduce the size of the corpus and identify the key words in the corpus as well as frequency distribution of concept words in overall context more precisely. After removing stop words, the tokenized sentence in above example contains:

[‘blockchain’, ‘supply’, ‘chain’, ‘match’, ‘made’, ‘heaven’]

In data pre-processing stage, we also conducted lemmatization using PorterStemmer of NLTK. The result shows that the corpus generated similar results in top30 most common words with or without stemming. We decided to adopt the result without stemming during to two reasons. First, some words may get over-stemmed (e.g. both ‘generous’ and ‘general’ are stemmed to ‘gener’) or under-stemmed (e.g. ‘bought’ remains ‘bought’ while ‘buy’ is stemmed to ‘bii’, while normally these two words have same stem ‘buy’). Second, some words cannot be correctly stemmed (e.g. ‘does’ is stemmed to ‘doe’).

To improve the accuracy, we have programmed to uniform the word forms to best possible extent. The measurements include to replace plural form with singular form (e.g. replace ‘systems’ with ‘system’) and to unify different expressions (e.g. replace ‘block chain’ with ‘blockchain’; ‘SCM’ with ‘supply chain Management’).

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

3.2 Training with Machine Learning Algorithm

Machine learning (ML) is the process of applying algorithms and statistical models to find patterns in massive amounts of data (MIT, 2018). With application of ML, computer systems can perform a specific task without having to relying on patterns or inferences. ML is seen as a subset of artificial intelligence (AI) and ML algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult to develop a conventional algorithm for effectively performing the task. This study applies Principal Component Analysis (PCA), word2vec and Latent Dirichlet Allocation (LDA) in Python environment to extract useful insights like dominant topics from the corpus.

With the recent advent of machine learning, topic modeling, in particular, providing an overview of themes being addressed in documents, has gained notable popularity as an innovative methodological approach, in a broad range of disciplines, including management and information systems (Hannigan et al., 2019). Topic modelling aims to reduce the complexity of compiling a large literature corpus by representing text as a combination of topics. Topics are clusters of words that reappear across texts, but the interpretation of these clusters as themes, frames, issues, or other latent concepts depends on the methodological and theoretical choices made by the researchers (Jacobs and Tschötschel, 2019).

As Hannigan et al., 2019 clearly illustrate, topic modeling’s technical and theory-building features are distinct from those of content analysis and general natural language processing of text. Topic modeling is a “rendering process,” for juxtaposing data and a problem domain theory, to generate new theoretical insights and artifacts such as technical and management constructs and the links between them. The topic modelling process, as a theory building tool, involves the rendering of corpora (preparing the sets of texts to be analyzed), the rendering of topics (making choices in view of how topics are identified within the text corpora), and the generation of theoretical artifacts (producing new theoretical constructs, identifying causal mechanisms and further insights out of the revealed topics).

3.2.1 Principal Component Analysis (PCA)

PCA is originally proposed by Hotelling (1933). It is a mathematical algorithm that derives the matrix of correlation coefficients from the original data set. The purpose of PCA is to reduce the dimensionalities of the data. In this study, PCA is carried out in four steps: (1) standardize data, (2) compute covariance matrix with the standardized data, (3) calculate eigenvectors and eigenvalues in the covariance matrix, and (4) sort the eigenvalues in decreasing order. The first principle component (PC1) will carry the most of variance and so on.

Efficiency of PCA model is measured by cumulative contribution rate (CCR). CCR is defined by the maximal amount of variance that is explained by the principal components representing the directions of the data. A CCR of 80% or above is considered acceptable in evaluating PCA model efficiency.

3.2.2 Word2vec

Mikolov et al. (2013) propose word2vec model for computing continuous vector representations of words from very large data sets, and observe large improvement in the quality of these representations measured in a word similarity. Word2vec is a two-layer neural network deep learning model that has a text corpus as input and the word vectors as output. To be more specific, it first constructs a dictionary of words from the training text data and then learns vector representation of those words.

The advantage is word2vec is that it detects similarities mathematically. It creates vectors that are distributed numerical representations of word features, such as the context of individual words. It then outputs a dictionary of words in which each word has a vector attached to it, which can be grouped to vectors of similar words or be fed into a deep learning model for further analysis. Through the algorithm word2vec establishes a word’s association with other words, which forms the basis of sentiment analysis and recommendations in various research domains.

Word2vec conducts semantic comparisons (Mikolov et al., 2013) ranging from country-currency (e.g. “India” is to “Rupee” as “Japan” is to “Yen”) and male-female (e.g. “man” is to “king” as “woman” is to “queen”).

Word2vec is applied in this study to assess word similarities, but the paper does not aim to discuss word2vec model in details. Interested party may refer to Rong (2014) which explains word2vec parameters in details.

3.2.3 Latent Dirichlet Allocation (LDA)

LDA is an unsupervised generative probabilistic method for modeling a textual corpus. It is used as a language model to cluster co-occurring words into topics. LDA builds a topic per document model and words per topic model, modeled as Dirichlet distributions (Blei, 2012). LDA assumes that each document can be represented as a probabilistic distribution over latent topics, and that the topic distributions in all documents share a common Dirichlet prior probability.

The basic idea of LDA is to compute the probability distribution over words. For a document in study, topics and their distributions in the text database are considered as latent variables or hidden structures. LDA model allows sets of observations to be explained by unobserved variables. When observations are words collected into documents, each document is a mixture of a small number of topics and each word's presence is attributable to one of the document's topics.

LDA is considered to be one of the most effective approaches (Blei et al., 2003) to model topics. Detailed explanation on LDA model is available in Blei et al. (2003), Steyvers and Griffiths (2007), and Blei (2012).

3.3 Visualization of Research Findings

3.3.1 Most Common Words

The most commonly used words in research data are reported in figure 2. The top 10 most common words are, 'blockchain', 'data', 'technology', 'system', 'information', 'management', 'service', 'transaction', 'business', 'model', respectively. This indicates that top concerns surrounding blockchain are relating to technology, information management, transaction and business model.

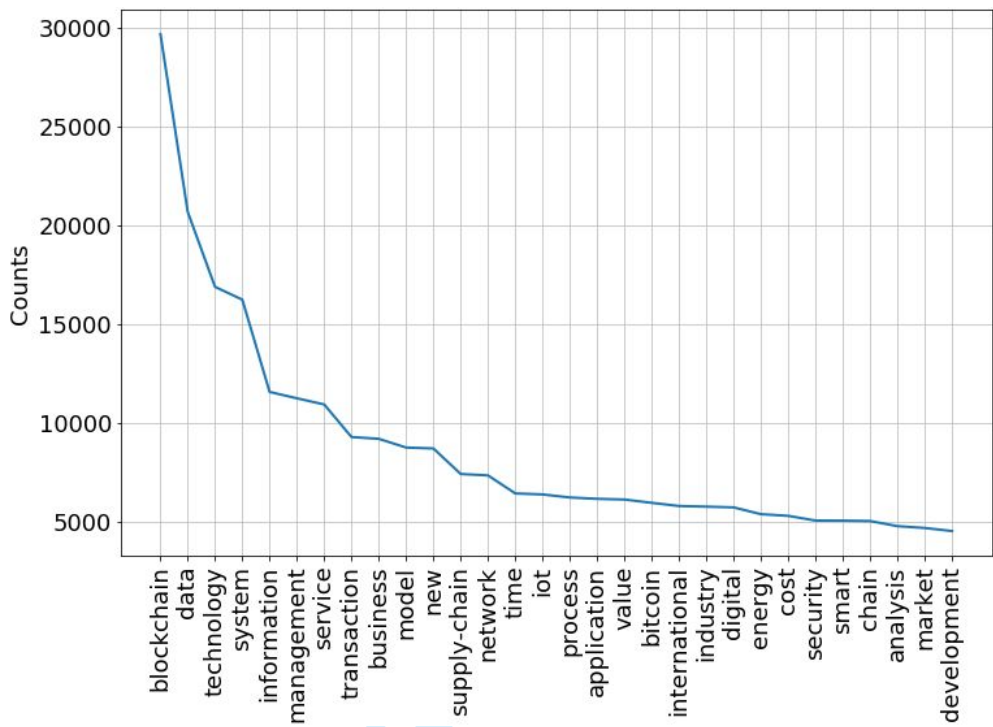


Figure 2 – Plot of most common words

3.3.2 Word Similarity

Word similarity is measured by cosine similarity, which is the cosine of the angle between two non-zero vectors of an inner product space. No similarity is expressed as a 90-degree angle, while total similarity of 1 is a 0-degree angle, complete overlap. Below is a list of words associated with “blockchain” using Word2vec, in order of proximity:

Table 2 – Word Similarity (top 20)

Word	Similarity	Word	Similarity
bitcoin	0.906903	analysis	0.693042
ledger	0.897149	challenges	0.683677
security	0.821995	privacy	0.674539
distributed	0.797730	trust	0.643097
block	0.795304	implementation	0.640840
application	0.756384	iot	0.637737
peer	0.754152	smart	0.630747
public	0.748685	access	0.628037
private	0.724296	technology	0.617000
design	0.715209	adoption	0.613373

4 Results and Discussions

4.1 Word Classification

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

By reducing the dimensions of word vectors by using PCA, most common words are classified (Figure 3). The plot illustrates three groups of key words:

- (1) 'blockchain', 'data', 'bitcoin', 'security' and 'application' are associated with each other closely;
- (2) 'business', 'industry', 'information', 'management' and 'system' are associated with each other closely;
- (3) 'international', 'cost', 'transaction' and 'energy' are outliers, which may be explained that these are popular topics, however, are not topics particularly related to supply chain blockchain.

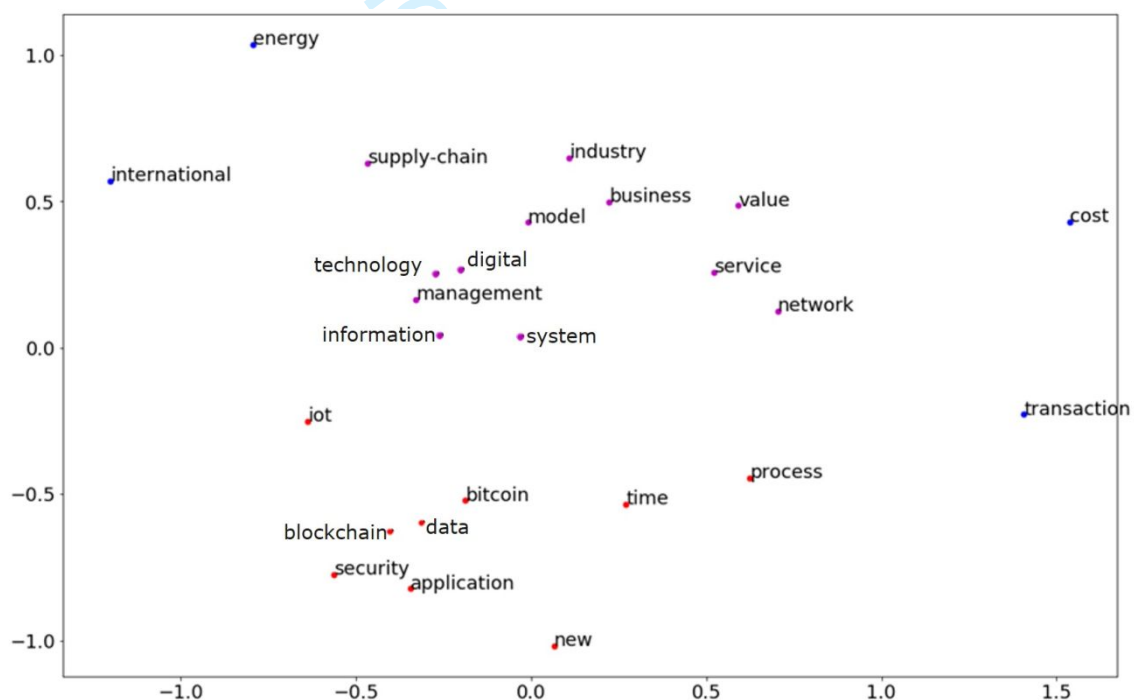


Figure 3 – Word classification

4.2 Dominant Topics

Applying to LDA model, four dominant topics are identified as listed in Table 3. In LDA models, each document is composed of multiple topics. But, typically only one of the topics is dominant. Setting the selection criteria as CCR ratio greater than 98%, four dominant topics are obtained. The CCR of dominant topic in relevant document is higher than 99.5%, which indicates the machine learning process was effective. The keywords in dominant topics concur with the top common words generated in 3.3.1.

Table 3 – Dominant topics

Topic_Num	Topic_Perc_Contrib	Keywords	Representative Text
0	0.0	0.99553 iot, system, datum, blockchain, base, network, security, device, service, application	[network, available, network, homepage, locate, challenge, way, forward, technology, internet, t...
1	1.0	0.99536 technology, supplychain, management, industry, datum, blockchain, system, business, service, model	[purchase, supply, management, available, purchase, supply, management, homepage, never, walk, a...
2	2.0	0.99672 blockchain, transaction, technology, system, information, process, base, platform, datum, business	[express, author, intend, represent, position, opinion, who, member, prejudice, member, obligati...
3	3.0	0.99678 supplychain, management, system, sustainability, service, performance, product, model, customer,...	[index, note, cross, refer, subentry, main, entry, main, entry, repeat, space, index, arrange, s...

To crosscheck against the top common words obtained in 3.3.1, we further plotted topics and weights (Figure 4). Notably ‘IoT’, ‘security’, ‘device’, ‘performance’, ‘customer’, ‘sustainability’ and ‘food’ are of higher weight (importance) regardless of lower frequency of appearance than other key words.

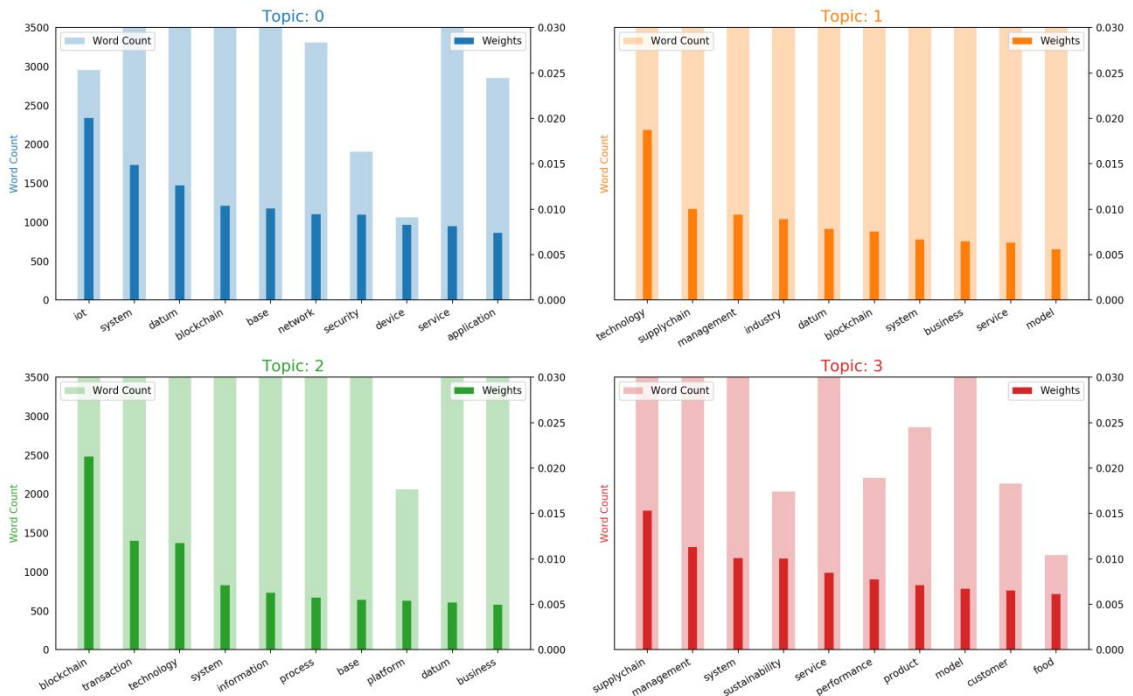


Figure 4 – Topics and Weights

The text mining and machine learning are performed via Python version 3.7.3 in macOS Catalina 10.15.3, MacBook Pro, Processor 2.4GHz Quad-Core Intel Core i5, Memory 16GB 2133 MHz LPDDR3. The computing time of each model is summarized in Table 4 as reference.

Table 4 – Computing time

Models	PCA	Word2vec	LDA
Computing time (seconds)	7.39088	181.52167	105.01739

5. Conclusion, research implications and future work

The research findings can be summarized in four perspectives. Firstly, for studies on blockchain application in supply chain, the top topics seem to be related to 'data', 'technology', 'system', 'information' and 'management' (Figure 2). Secondly, blockchain is considered of higher similarity to 'bitcoin', 'distributed ledger', 'security' and 'application' (Table 2). Thirdly, 'IoT', 'security', 'device', 'performance', 'customer', 'sustainability' and 'food' are of higher weight (importance) regardless of lower frequency of appearance than other key words (Figure 4). Fourthly, 'design', 'trust', 'implementation', 'challenges', and integration with 'IoT' are of higher concern than other perspectives like 'standardization', 'interoperability' and 'regulation' etc. (Table 2, Table 3).

The research insights and implications, as derived from our study, are three-fold. Firstly, while predominantly, generic technology and management challenges are of concern, focus should be given to particular design and implementation aspects of blockchain in the supply chain field. Blockchain deployments in practices are mostly in the pilot stage as of yet (Queiroz et al., 2019). Future focus should be given to develop architecture of blockchain solutions to provide seamless network and transparency in supply chains that benefits public safety and security.

Secondly, integration with IoT is considered to be of high importance. IoT has been rapidly applied in various areas of supply chain management in the past two years, security of information is of paramount issue. Blockchain technology has been explored as one option to effectively address those security concerns, allowing the advantage of decentralized data management. Kshetri (2017) suggests that the integration of IoT data to a blockchain platform could potentially further improve the overall efficiency. Blockchain can play a key role in tracking the sources of vulnerability in supply chains and in handling crisis situations like product recalls that occur after safety and security vulnerabilities are found. The IoT ecosystem is evolving quickly, developing several applications in different sectors. As such, future research agenda may be set to explore secure technical solutions to integrate IoT in supply chains in the context of increasing malicious IoT treats. Regulation of IoT security and data protection need to be developed and strengthened.

Thirdly, blockchain plays a crucial role in food sustainability. The technology could help consumers and businesses understand whether their products were produced sustainably and avoid environmentally damaging, illegal or unethical products. The blockchain-based supply chain traceability and transparency technology helps drive increased responsible production and consumption. New technologies such as IoT and blockchain can accelerate the progress of supply chain sustainability. How quick these technologies are adopted and implemented is becoming a key to protect environment and relax pressure on food shortage. Modern supply chains are complex and require digital connectivity and agility across participants, business leaders need to understand what needs to change in their organization to leverage blockchain implementation effectively.

Unlike previous researches that have mainly applied analysis to bibliometric data and/or abstract, this paper analyzes full text contents of paper with machine learning models to generate insights.

Due to access constraints, only scientific papers published in Science Direct, Springer and Emerald are included in this study. This may potentially bias the research findings. Future studies may consider applying a larger size of corpus. In addition, it could also be of value to train with different set of machine learning models.

To the best of our knowledge, our research is the very first attempt to apply machine learning to blockchain related research in the supply chain sector, thereby providing new insights and complementing existing literature.

References

Ar, I. M., Erol, I., Peker, I., Ozdemir, A. I., Medeni, T. D., and Medeni, I. T. (2020), "Evaluating the Feasibility of Blockchain in Logistics Operations: A Decision Framework", *Expert Systems with Applications*, Vol. 158, No. 113543.

Asmussen, C. B., and Charles. M. (2019), "Smart Literature Review: A Practical Topic Modelling Approach to Do Exploratory Literature Review", *Journal of Big Data*, Vol. 6, Article No. 93

Asmussen, C. B., and Charles M. (2020), "Enabling supply chain analytics for enterprise information systems: a topic modelling literature review and future research agenda", *Enterprise Information Systems*, Vol. 14, No. 5, pp. 563-610.

Bai, C., and Sarkis, J. (2020), "A supply chain transparency and sustainability technology appraisal model for blockchain technology", *International Journal of Production Research*, Vol. 58, No. 7, pp. 2142-2162.

Blei, D.M., Ng, A.Y, and Jordan, MI (2003), "Latent Dirichlet Allocation", *Journal of machine Learning research*, Vol. 3, pp.993-1022.

Blei, D. M. (2012), "Probabilistic topic models", *Communications of the ACM*, Vol. 55, No. 4, pp. 77-84.

Blossey, G; Eisenhardt, J; and Hahn, G. (2019), "Blockchain technology in supply chain management: An application perspective", *In Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

Casey, M.J. and Wong, P. (2017), "Global supply chains are about to get better, thanks to blockchain", *Harvard Business Review*, Vol. 13, pp. 1-6.

Chang, S.E. and Chen Y. (2020), "When blockchain meets supply chain: A systematic literature review on current development and potential applications", *IEEE Access*, Vol. 8, pp.62478-62494.

- Duru O. and Zin M.I.M. (2019), Blockchain Roaming in the Maritime Industry, <https://splash247.com/blockchain-roaming-in-the-maritime-industry/>, Date of access: 27/06/2020.
- Eickhoff, M. and Neuss, N. (2017), "Topic Modelling Methodology: Its Use in Information Systems and Other Managerial Disciplines", *In Proceedings of the 25th European Conference on Information Systems (ECIS), Guimarães, Portugal, June 5-10, 2017*, pp. 1327-1347.
- Fiskin, C.S. and Cerit, A.G. (2020), "Comparative bibliometric and network analysis of maritime transport / shipping literature using the Web of Science literature", *Scientific Journals of the Maritime University of Szczecin*, Vol. 61, No. 133, pp. 160-170.
- Griffiths, T. L., Steyvers, M., and Tenenbaum, J. B. (2007), "Topics in semantic representation", *Psychological review*, Vol. 114, No. 2, pp. 211-244.
- Gurtu, A. and Johny, J. (2019), "Potential of blockchain technology in supply chain management: A literature review". *International Journal of Physical Distribution & Logistics Management*, Vol. 49, No. 9, pp. 881-900.
- Hannigan, T.R., Haans, R.F., Vakili, K., Tchalian, H., Glaser, V.L., Wang, M.S., Kaplan, S., and Jennings, P.D. (2019), "Topic Modeling in Management Research: Rendering New Theory from Textual Data", *The Academy of Management Annals*, vol. 13, pp. 586-632.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology*, Vol. 24, No. 6, pp. 417-441.
- Iansiti, M., and Lakhani, K. R. (2017), "The truth about blockchain", *Harvard Business Review*, Vol. 95, No. 1, pp. 118-127.
- Jacobs, T., and Tschötschel, R. (2019), "Topic models meet discourse analysis: A quantitative tool for a qualitative approach", *International Journal of Social Research Methodology*, Vol. 22, No. 5, pp. 469-485.
- Kshetri, N. (2017), "Can blockchain strengthen the internet of things?", *IT professional*, Vol.19, No. 4, pp. 68-72.
- Kshetri, N. (2018), "Blockchain's roles in meeting key supply chain management objectives", *International Journal of Information Management*, Vol. 39, pp. 80-89.
- Kummer, S., Herold, D.M., Dobrovnik, M., Mikl, J., Schäfer, N. (2020), "A Systematic Review of Blockchain Literature in Logistics and Supply Chain Management: Identifying Research Questions and Future Directions", *Future Internet*, Vol. 12, No.3-60, pp. 1-15.
- Lambrou, M., Watanabe, D. and Iida, J., (2019a), "Maritime Blockchains: Decoding Diverse Strategies for Value Extraction", *27th Annual Conference of the International Association of Maritime Economists (IAME)*, Paper ID 114, 1-22.
- Lambrou, M., Watanabe, D. and Iida, J., (2019b), "Shipping digitalization management: conceptualization, typology and antecedents", *Journal of Shipping and Trade*, Vol. 4, Article No. 11.

- Lee, S.W., and Shin S.H. (2019), “A review of port research using computational text analysis: A comparison of Korean and International Journal”, *The Asian Journal of Shipping and Logistics*, Vol. 35, No. 3, pp.138-146.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013), “Efficient estimation of word representations in vector space”, *arXiv preprint arXiv:1301.3781*.
- MIT (2018), <https://www.technologyreview.com/2018/11/10/139137/is-this-ai-we-drew-you-a-flowchart-to-work-it-out/>, Date of access: 02/10/2020.
- Pournader, M., Shi, Y., Seuring, S., and Koh, S. L. (2020), “Blockchain applications in supply chains, transport and logistics: a systematic review of the literature”, *International Journal of Production Research*, Vol. 58, No. 7, pp. 2063-2081.
- Queiroz, M. M., Telles, R., and Bonilla, S. H. (2019), “Blockchain and supply chain management integration: A systematic review of the literature”, *Supply Chain Management: An International Journal*, Vol. 25, No. 2, pp. 241-254.
- Risius, M., and Spohrer, K. (2017), “A Blockchain Research Framework”, *Business and Information Systems Engineering*, Vol. 59, No. 6, pp. 385-409.
- Roeck, D., Sternberg, H., and Hofmann, E. (2020), “Distributed ledger technology in supply chains: A transaction cost perspective”, *International Journal of Production Research*, Vol. 58, No. 7, pp. 2124-2141.
- Rogerson, M., and Parry, G. C. (2020), “Blockchain: case studies in food supply chain visibility”, *Supply Chain Management*, Vol. 25, No. 5, pp. 601-614.
- Rong, X. (2014), “word2vec parameter learning explained”. *arXiv preprint arXiv:1411.2738*.
- Saberi, S., Kouhizadeh, M., Sarkis, J., and Shen, L. (2019), “Blockchain technology and its relationships to sustainable supply chain management”, *International Journal of Production Research*, Vol. 57, No. 7, pp. 2117-2135.
- SAS. (2020), https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html. Date of access: 27/06/2020.
- Shahid, M. N. (2020), “A Cross-Disciplinary Review of Blockchain Research Trends and Methodologies: Topic Modeling Approach”, In Proceedings of the 53rd Hawaii International Conference on System Sciences.
- Shin, S.H., Kwon, O.K., Ruan, X., Chhetri, P., Lee, P.T.W., Shahparvari, S. (2018), “Analyzing sustainability literature in maritime studies with text mining”, *Sustainability*, Vol. 10, No. 10-3522, pp. 1-19.
- Sternberg, H., and Baruffaldi, G. (2018), “Chains in chains: Logic and Challenges of Blockchains in Supply Chains”, In *51st Hawaii International Conference on System Sciences*.
- Steyvers, M., and Griffiths, T. (2007), Probabilistic topic models. *Handbook of latent semantic analysis*. New Jersey: Lawrence Erlbaum Associates.
- Treiblmaier, H. (2018), "The impact of the blockchain on the supply chain: a theory-based research framework and a call for action", *Supply Chain Management*, Vol. 23 No. 6, pp. 545-559.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Tsiulin, S., Reinau, K. H., Hilmola, O. P., Goryaev, N., and Karam, A. (2020), "Blockchain-based applications in shipping and port management: a literature review towards defining key conceptual frameworks", *Review of International Business and Strategy*, Vol. 30, No. 2, pp. 201-224.
- Wagner, N., and Wiśnicki, B. (2019), "Application Of Blockchain Technology In Maritime Logistics", *In Diem: Dubrovnik International Economic Meeting*, Vol. 4, No. 1, pp. 155-164.
- Wang, Y., Han, J. H., and Beynon-Davies, P. (2019), "Understanding blockchain technology for future supply chains: a systematic literature review and research agenda", *Supply Chain Management: An International Journal*, Vol 24. No. 1, pp. 62-84.
- Westergaard, D., Størfeldt, 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, pp. e1005962, pp. 1-16.