Received 15 February 2016
Revised 14 June 2016
9 August 2016
1 September 2016
Accepted 2 September 2016

The data-driven analytics for investigating cargo loss in logistics systems

Pei-Ju Wu
Department of Transportation Technology and Management, Feng Chia University, Taichung, Taiwan
Mu-Chen Chen
Department of Transportation and Logistics Management, National Chiao Tung University, Taipei, Taiwan, and
Chih-Kai Tsau
Degree Program of Transportation and Logistics, National Chiao Tung University, Taipei, Taiwan

Abstract

Purpose – Cargo loss has been a major issue in logistics management. However, few studies have tackled the issue of cargo loss severity via business analytics. Hence, the purpose of this paper is to provide guidance about how to retrieve valuable information from logistics data and to develop cargo loss mitigation strategies for logistics risk management.

Design/methodology/approach – This study proposes a research design of business analytics to scrutinize the causes of cargo loss severity.

Findings – The empirical results of the decision tree analytics reveal that transit types, product categories, and shipping destinations are key factors behind cargo loss severity. Furthermore, strategies for cargo loss prevention were developed.

Research limitations/implications – The proposed framework of cargo loss analytics provides a research foundation for logistics risk management.

Practical implications – Companies with logistics data can utilize the proposed business analytics to identify cargo loss factors, while companies without logistics data can employ the proposed cargo loss mitigation strategies in their logistics systems.

Originality/value – This pioneer empirical study scrutinizes the critical cargo loss issues of cargo damage, cargo theft, and cargo liability insurance through exploiting real cargo loss data.

Keywords Decision tree, Cargo loss, Data-driven analytics, Logistics risk management, Logistics system

Paper type Research paper

Introduction

Risk management has been a crucial issue in logistics systems for decades. Identifying and assessing risk is an essential process to implement supply chain risk management (Kern *et al.*, 2012). Hedging, security, visibility, avoidance, speculation, and postponement, as well as communicative and cooperative relationships, are general strategies for supply chain risk management (Manuj and Mentzer, 2008; Wieland and Wallenburg, 2013; Durach *et al.*, 2015; Hohenstein *et al.*, 2015; Hale and Moberg, 2005; Chopra and Sodhi, 2014). In fact, numerous studies have macroscopically probed logistics safety and supply chain risk management (Vilko and Hallikas, 2012). Herein, cargo loss is a critical problem among the risk items of a logistics process. However, rarely do studies provide empirical investigations focusing on

This work is partially supported by Ministry of Science and Technology, Taiwan, ROC under Grant No. MOST 103-2410-H-009-029-MY3. The authors also wish to thank the Editor-in-Chief and referees for their helpful comments and suggestions. Any errors or omissions remain the sole responsibility of the authors.

0960-0035
DOI 10.1108/IJPDLM-02-2016-0061

identifying and evaluating cargo loss in logistics systems or furnish corresponding mitigation strategies for risk management.

Once cargo loss occurs, the most direct consequence is the financial loss borne by the cargo owner. Most firms purchase cargo transport insurance, enabling them to claim substantial monetary compensation. However, the insurance amount expended increases as the monetary amount of losses incurred rises. Therefore, if firms merely rely on cargo insurance, the accumulated amount of insurance claims will ultimately be reflected in their future insurance costs. The financial loss entails not only the cost required for reproducing the products, but also fees arising from handling the loss incident, increased insurance costs, loss of business market opportunities, and the negative impacts on the firm's reputation (Burges, 2012). Hence, cargo loss in logistics systems is a topic warranting immediate attention. Nevertheless, few studies have investigated how to mitigate cargo loss in the field of logistics risk management by exploring logistics big data.

This study aims to tackle the causality of cargo loss severity in logistics systems through big data analytics. Specifically, this study proposes a framework of business analytics involving descriptive, predictive, and prescriptive analytics to extract essential patterns of cargo loss incidents of the case company. Such data-driven approaches can retrieve potentially valuable information from data and conquer real-world problems (Archak *et al.*, 2011; Zanakis and Becerra-Fernandez, 2005; Schoenherr and Speier-Pero, 2015). The empirical results of the data-driven analytics demonstrate that transit types, product categories, and shipping destinations are essential factors determining cargo loss severity. Then, the cargo prevention strategies are further derived from the empirical results for risk management in logistics systems.

The main contributions of this study can be summarized as follows: to the best of our knowledge, this paper is the first research to employ real data of cargo loss to explore the hidden sides of logistics systems. Specially, most companies own substantial data but face challenges regarding how to exploit their data. Hence, based on the proposed cargo loss business analytics, managers can identify which logistics circumstances are more likely to trigger cargo loss incidents, and subsequently allocate resources to prevent cargo loss in their logistics systems. This pioneer empirical study provides clear guidance about how to turn logistics data into valuable insight. Moreover, companies without logistics data (e.g. new entrants, from local to global logistics) can apply the proposed mitigation strategies for logistics risk management derived from this empirical study to their logistics systems. Furthermore, this study scrutinizes the critical cargo loss issues of cargo damage, cargo theft, and cargo insurance, which makes a valuable contribution to the practice of logistics risk management.

The rest of this paper is organized as follows. Second section reviews the existing literature. Third section proposes a research design of business analytics for cargo loss. Fourth section then analyzes a case of shipping precise electronic products. Next, fifth section provides essential risk management strategies for cargo loss prevention. Finally, sixth section draws conclusions.

Literature review

A literature review of studies related to cargo loss, cargo liability insurance, and data-driven analytics has been conducted.

Cargo loss

In general, logistics systems may encounter two forms of cargo loss involving cargo damage and cargo theft. Cargo damage and cargo theft both cause firms severe financial loss and result in the risk of not delivering on time to their customers. Here, the former indicates that cargos suffer damage, and the latter denotes that cargos disappear in a planned logistics process.

Cargo damage

Kokotos and Smirlis (2005) investigated the risk events ensuing in a logistics process, such as accidents on ships at sea. Kutz (2007) examined product packaging methods to discuss how appropriate packaging can be used to effectively reduce the number of cargo loss incidents. Although increasing packaging protection decreases damage and theft in transit, it increases package weight and cost, thereby elevating transportation costs (Lambert *et al.*, 1998). Nevertheless, few studies have explored valuable cargo loss data from insurance reimbursement claim databases to deal with the causes of cargo loss incidents in logistics systems.

Cargo theft

Among cargo loss-related topics, cargo theft has garnered the attention of academic experts and practitioners. Planned cargo theft by criminal groups all over the world is one of the most common cause of severe financial loss; specifically in recent years, electronic products have become the target most preferred by thieves worldwide (FreightWatch, 2013). FreightWatch (2013) mentioned that regions exhibiting a high risk for theft in 2013 were Mexico, Brazil, South Africa, the USA, and Russia. Moreover, the number of thefts in European countries has also increased annually, with electronic products as the most preferred target (FreightWatch, 2013).

Several studies centering on cargo loss issues have investigated cargo thefts worldwide and how such thefts can be avoided (Burges, 2012). Ekwall (2009) utilized crime displacement theory to investigate the phenomenon of cargo theft in the transport system. Burges (2012) found that a single cargo theft incurs a total supply chain cost six times the value of the cargo itself and further classified the impacts of cargo theft into the following categories: costs of product replacement, costs of handling a theft incident, increased insurance premium, loss of sales, and negative impact on the business' reputation. The Transported Asset Protection Association (2012) pointed out that cargo thefts use the following fraudulent approaches to steal cargos: the creation of a fake transportation company, impersonation of real transport companies, fraudulent pickups. Cargo loss prevention is a persistent problem that is particularly crucial when the cargo involves high-priced, fragile products that can be easily stolen, and for which all logistics operations must therefore be performed according to a standard operating procedure. For example, cargo delivery time must be predefined so that when the cargo reaches its destination, it can be collected immediately, eliminating the need to temporarily store expensive products at a location exposing them to the risk of being stolen (Fennelly, 2012). Additionally, Ekwall and Lantz (2013) explored the seasonality of cargo theft in European, Middle-Eastern, and African counties.

Cargo liability insurance

Most studies have applied cargo insurance concepts and international cases of cargo insurance claims to inspect cargo loss incidents based on the viewpoints of the cargo insurance sector (Skorna and Fleisch, 2012). For example, the transport underwriting systematic risk can be appraised from the viewpoint of insurance financial pricing (Lai, 2008).

In practice, cargo loss has been extensively investigated by insurance companies, with the aim of reducing cargo loss incidents and insurance compensation. Specifically, large international insurance companies often invite cargo loss prevention teams and experts to conduct relevant risks analyses. The International Group of Protection and Indemnity Club (P&I Club) founded by ship owners from various countries has endeavored to prevent cargo loss occurrences (UK P&I Club, 2013). Moreover, the world's largest P&I Club, UK P&I Club, established the Carefully to Carry Committee in 1961, affording a wide variety of cargo loss prevention information and experiences (UK P&I Club, 2013). In 1997, Intel, Sony, Hewlett Packard, and Dell jointly formed the Transported Asset Protection Association (TAPA), which focuses on formulating safety standard regulations for logistics operators in the supply chain of high-tech industries (Transported Asset Protection Association, 2012). Furthermore, these regulations are safety certifications for transportation or cargo storage operators to which numerous technology industries seek to outsource these services, because one of the risk items managed by TAPA is preventing high-tech products in every stage of the logistics process from being stolen (Transported Asset Protection Association, 2012).

Global logistics involves various actors with different liabilities to manage cargo loss risk. Nonetheless, cargo loss responsibilities are not easy to identify. Accordingly, cargo owners and logistics service providers resort to distinct types of insurance policies. Cargo owners, especially international brand companies needing frequent global logistics, are inclined to purchase Stock Throughput (STP) insurance. STP insurance covers all static and dynamic logistics of cargo owners (Cameron and Sheridan, 2013). Hence, cargo owners do not need to have cargos insured each time whenever each logistics activity occurs. On the other hand, logistics service providers usually take out liability insurance to mitigate cargo loss risk. That is, once cargo owners or cargo owner's insurance companies exercise the right of subrogation on logistics service providers, liability insurance can be triggered to compensate cargo owners for financial loss.

Data-driven analytics

In the environment of intense competition, various levels of business decisions can be effectively made based on more data and more analysis (Davis, 2014). Moreover, the decision-making capability of businesses can be developed through implementing business analytics.

Regarding recent real-world applications of business analytics, Ford utilized innovative analytics for the program called the Ford Fleet Purchase Planner that benefits both customers and the environment. (Reich *et al.*, 2015). iHeartMedia, a large media company having over 850 stations in more than 150 cities, first investigated the need for analytics to encounter the challenges it faced, and then utilized business analytics to schedule radio advertisements of stations (Venkatachalam *et al.*, 2015). Ingram Micro, the largest electronics distributor in the world, has created enormous profit through utilizing business analytics to discover sales opportunities and to implement data-driven marketing campaigns (Mookherjee *et al.*, 2016).

Summary

A growing body of literature investigates the issues of cargo damage, cargo theft, and cargo liability insurance in cargo loss problems. The studies on cargo damage focus on shipping accidents (Kokotos and Smirlis, 2005), packaging methods to diminish cargo loss (Kutz, 2007), costs of packaging protection (Lambert *et al.*, 1998); the literature on cargo theft emphasizes hedges against cargo theft (Burges, 2012; Fennelly, 2012) and the behavior of cargo theft (FreightWatch, 2013; Ekwall, 2009; Transported Asset Protection Association, 2012; Ekwall and Lantz, 2013); the literature on cargo liability insurance stresses cargo loss prevention and insurance compensation (UK P&I Club, 2013; Transported Asset Protection Association, 2012), cargo loss liability and insurance (Cameron and Sheridan, 2013), cargo insurance claims (Skorna and Fleisch, 2012), and transport underwriting risk (Lai, 2008).

Based on the above literature review, few studies have scrutinized real cargo loss data to tackle the causes of cargo loss incidents with emphasis on logistics systems. To fill this research gap, this study devises cargo loss business analytics to systematically discover the critical cargo loss issues of cargo damage, cargo theft, and cargo liability insurance from an insurance reimbursement claim database as well as to holistically create the corresponding strategies for cargo loss prevention in logistics systems building on the extant literature.

Business analytics for cargo loss severity

This study devised a research design incorporating the knowledge discovery in databases (KDD) procedure (Fayyad and Stolorz, 1997; Figueiredo *et al.*, 2005; Roiger and Geatz, 2003) and business analytics of descriptive, predictive, and prescriptive analyses (Evans, 2012; Souza, 2014; Phillips-Wren and Hoskisson, 2015) to tackle cargo loss issues in logistics systems. The proposed research design of business analytics for cargo loss severity (Figure 1) is briefly described as follows.

Descriptive analytics

- Data collection and exploration: this study gathered comprehensive cargo loss data from the case company, and conducted descriptive statistics on them to understand their basic characteristics.
- Data selection: the data that are unrelated to cargo loss events were removed to ensure that the chosen data can be used to develop cargo loss models.
- Data preprocessing: data preprocessing involves data integration and data cleaning. Data integration aims to remove inconsistent data and deletes duplicate data. Inconsistent data are mainly caused by different recording methods. For instance, the same country may be recorded with different abbreviations; the same product may have varying product descriptions; the same freight forwarder may have inconsistent company names; the same amount of cargo loss may be recorded in different currencies in different countries. These inconsistencies in data should be modified at this stage to ensure that research data are standardized. Meanwhile, data cleaning ensures that data are accurate and intact. The main task in assessing data accuracy is to determine whether attribute values are valid or located within the appropriate range. Cross-inspection is an effective approach for assessing the accuracy of data.

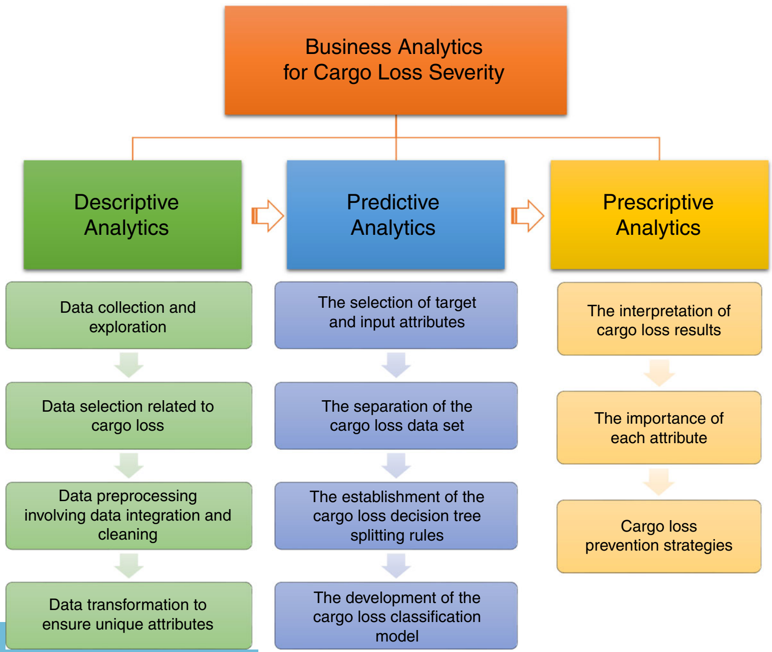

Figure 1.
The proposed business analytics for cargo loss severity

For example, an inaccuracy would be detected if a real loss amount is larger than an invoice amount. When incorrect attribute values are encountered, the entire instance will be deleted. To ensure intact data, missing data were checked for. If missing values could be inferred based on other attribute values, they were added; otherwise, the entire data entry was deleted.

- Data transformation: since data are recorded in various formats, approaches for data transformation include normalizing data, adding or deleting attributes, and transforming data attribute types (Roiger and Geatz, 2003). To create new attributes, this study re-classified data that are too trivial. For example, date values were transformed into month data values; country names were transformed into region names; and various freight forwarders or truck companies were transformed into major types based on logistics service methods. Furthermore, this study transformed continuous values into categorical ones. For instance, cargo loss amounts were classified into different loss levels to reflect cargo loss severity.

Predictive analytics

Data mining is an essential method in predictive analytics. After rigorously formalizing gross data, this study utilized decision tree analysis in data mining to generate a cargo loss model. The primary process to establish a decision tree model is as follows:

- Select target and input attributes: based on the research purpose, the target attribute of cargo loss severity was selected from all attributes. Then, the attributes related to the target attributes were chosen as input attributes to construct the classification model.
- Separate the data set: before developing a data mining model, the chosen data set should be divided into training data, validation data, and test data.
- The establishment of the decision tree splitting rules: the split parameters of the cargo loss decision tree, such as split rule, maximum branch, and maximum tree depth, should be determined. This study adopted default values of SAS Enterprise Miner. Based on default value results, this study examined and adjusted the parameters to ensure the accuracy of classification and meaningful managerial implications.
- The development of the cargo loss classification model: this step generates the classification model of cargo loss severity based on domain knowledge of logistics operations and the accuracy rate of training data, validation data, and test data. Moreover, an importance ranking of data attributes and decision rules can be established.

Prescriptive analytics

The main component of prescriptive analytics is to evaluate and interpret cargo loss classification results from decision tree analysis, including the accuracy rate of the decision tree classification, the coverage rate of leaf nodes, and the priority rank of input variables for the splitting decision tree. Examining the accurate classification rate of validation data and test data as well as the coverage rate of leaf nodes can ensure the predictive ability of classification results by the decision tree. Furthermore, the importance of each input variable to cargo loss severity can be determined by evaluating the spilt priority of input variables on the decision tree. Additionally, the split and span of the decision tree should be carefully analyzed from managerial perspectives. Thereafter, managers can develop their prevention strategies to avoid cargo loss.

A case of electronic products

The case background

The case company is a brand electronics manufacturer for global logistics covering the consumer electronics market around the world. The expensive electronic products (e.g. computers, smart phones, etc.) sold by the case company are very vulnerable to damage by shaking or crushing.

The case company has large-scale business with a competitive global logistics network involving air freight forwarders, sea freight forwarders, truck companies, and global express. Hundreds of flights are shipped around the world every day. Due to such frequent and complicated intermodal shipments, the case company must deal with cargo loss incidents every day worldwide. The only way that the case company currently handles cargo loss issues is to purchase global cargo insurance as its main form of logistics risk management. Nonetheless, cargo insurance is not the fundamental way to prevent cargo loss. Furthermore, as the rate of cargo loss increases each year, the case company must then pay more cargo insurance fees for the next year. Consequently, this study aims to clarify the case company's cargo loss situation and to create corresponding strategies for prevention of cargo loss.

Results

The collected cargo loss data were analyzed by the proposed business analytics for cargo loss severity. The results of the business analytics can be summarized as follows.

Descriptive analytics

Data collection and exploration. This study collected cargo loss data from the insurance reimbursement claim database of the case company's insurance company covering 1,416 cargo loss cases in its logistics system in 2011. This implies that nearly four cargo loss cases happened per day worldwide. Additionally, it should be noted that the data set does not seem to be as large as big data issues, but actually the set of raw data of the case company is enormous, and it cannot be presented in the paper due to confidentiality rules. Nevertheless, the proposed business analytics is developed based on the procedure of big data analytics, and it can be utilized to tackle big data issues. The attributes and definitions of the collected data are summarized in Table I.

After reviewing the cargo loss data collected from the case company, this study conducted descriptive statistics to briefly understand the case company's cargo loss data. Regarding the proportion of cargo loss by transit type, cargo loss (claim payments) experienced by the case company had occurred during air, sea, and land transportation. As shown in Table II, the primary cargo loss in terms of monetary value was associated with sea transport, followed by air transport and truck transport. However, the frequency of cargo loss cases involving air transport was around three times that of sea transport. Concerning the proportion of cargo loss by product category, laptops dominated the cargo loss not only in financial loss but also in the frequency of loss cases. Since Table II merely identifies various levels of the frequency and claims payments of cargo loss in terms of transit type and product category, the impact of the numerous combinations of logistics risk factors on different levels of cargo loss severity should be further investigated by a rigorous method.

Data selection. Field attributes unrelated to the cargo loss issues were removed. Accordingly, applicant, master waybill number, house waybill number, and invoice number were eliminated.

Data preprocessing. Data integration involved removing inconsistent data and repetitive data; data cleaning was aimed at ensuring data accuracy and integrity. Table III summarizes how this study performed data integration and cleaning. In data preprocessing, from the 13 data fields present in the 1,416 pieces of data for 2011, five field attributes were

Attributes	Definitions
Applicant	Applicants involved in cargo loss
On board date	The date each cargo was loaded on board the transport vessel
Reporting date	The date that case company claimed reimbursement
Master waybill no.	The number records cargo information of air waybill or marine bill of lading for air and sea transportation, respectively; consignment bill or packing list for land transportation
House waybill no.	The number records sub-bill of lading for air and sea transportation; house waybill no. of land transportation is the same as master waybill no.
Shipper	The unit which sent cargo. In this study, the shipper is all the case company
Consignee	The unit which received cargos involving retailers and wholesalers
Product category	Product category that suffered cargo loss
Ship from	The country that the cargo shipped from
Ship to	The country that the cargo shipped to
Unit price	Product price as shown on the invoice issued by the seller
Loss quantity	Quantity of cargo loss recorded
Loss amount	The total monetary loss calculated based on cargo unit price and the actual quantity of cargo loss
Claim payment	The monetary amount of an insurance claim should equal the total monetary cargo loss; in a specific dispute, if the insurance company refused to pay the full claim amount, the actual compensation amount may be smaller than the total monetary cargo loss
Total invoice value	The total monetary amount for the cargo involved in all stages of the logistics operation as recorded on the seller-issued invoice
Invoice no.	All invoice numbers for the cargo involved in all stages of the logistics operation
Transit type	Transit type includes air, sea, and land transports. The specific type of intermodal transportation used when cargo loss occurred and the stage of logistics operation are recorded
Transporter	Transporters include sea/air/land transport carriers and express delivery forwarders

Table I. The attributes and definitions of cargo loss data

Attributes	Claim payments (%)	Frequency of loss cases (%)
Transit type		
Air	13.17	59.31
Sea	77.20	21.76
Truck	9.63	18.93
Total	100.00	100.00
Product category		
Desktop	2.60	0.54
LCD monitor	3.29	1.69
MB/VGA cards	20.70	7.20
Laptop	38.81	83.83
Others	29.13	3.14
Pad	5.34	2.76
Wireless LAN	0.13	0.84
Total	100.00	100.00

Table II. The cargo loss by transit type and product category

removed, namely reporting date, unit price, loss quantity, loss amount, and total invoice value. In addition, missing or typographical error (typo) data observation values were deleted from 111 data fields, converging the data set into eight data fields and 1,305 cargo loss observation values.

Data transformation. Two category attributes' values (on board date and product category) were converged, and three new category attributes (ship to area, forwarder type, and financial impact) were created during data transformation. Table IV summarizes how data transformation was performed.

Table III.
Data preprocessing

Insurance claim report data fields	Data integration	Data cleaning
On board date	Date format was checked and unified	Whether the on board date was earlier than the reporting date was checked. Data containing typos or missing values were eliminated. Since the case company's reporting date differed from the time at which cargo loss was discovered by several weeks, this field attribute was deleted
Reporting date	Date format was checked and unified	
Product category	Product name was checked and unified	Data containing typos and missing values were eliminated
Ship from	Country name was checked and unified	Cross-validation was used to determine whether typos pertaining to departure and destination countries were present. Supplementary information was provided if departure and destination countries were identifiable; if not, such data were deleted
Ship to	Country name was checked and unified	
Unit price Loss quantity	Unit price was multiplied by loss quantity to yield total loss amount. Loss amount was retained to express the actual loss, and the unit price and loss quantity fields were deleted	The accuracy of the field values was checked by determining whether unit price multiplied by loss quantity equaled loss amount and claim payment. Data containing erroneous field values or missing values were deleted
Loss amount Claim payment	Values under loss amount and claim payment were identical. Claim payment was retained, deleting the loss amount field	
Total invoice value	Check and unify currency	Data with missing values were eliminated. The total invoice values recorded were incorrect and therefore this attribute was deleted
Transit type	Field values were checked and unified; only three transit types (land, air, and sea) were used	Cross-validation was performed on departure and destination countries to avoid typos. Data containing typos were corrected if the departure and destination countries were identifiable; if not, such data were deleted entirely
Transporter	Company name of the transporter was checked and unified	Field error values or missing values were corrected or supplemented according to the transit route commonly used by logistics companies

Herein, the cargo loss severity, in practice, is usually evaluated by the amount allowable for loss (claim payment). Consequently, financial impact is used to represent cargo loss severity. Furthermore, three levels of cargo loss severity can be classified based on practical management as follows:

- Low severity: a loss amount of less than US$200 is directly compensated by the insurance company. The case company perceives this loss as exerting minimal financial influence and thus defined the level of severity for this loss as "low."
- Medium severity: for a loss amount of US$201-US$5,000, the insurance company must conduct extensive investigations according to stringent claim criteria. The level of severity for this degree of loss is defined as "medium."
- High severity: a loss amount of US$5,001 and higher is perceived by the insurance company to be a large sum. The insurance company will delegate notary agencies to conduct onsite investigation, a process that the case company considers to be time

Insurance claim report data fields	Before transformation	After transformation
On board date	This information was expressed in terms of "day"	This information was converged to "month"
Product category	This information recorded the categories of products involved in loss incidents; 14 product categories were identified	Products involved in less than three cargo loss incidents throughout the entire year were merged with the "Others"
Ship to	This information was expressed in terms of country names	A new category attribute (ship to area) was created by consolidating the destination country names into regions: Asia, Europe, North America, South America, Australia, Africa, and Middle-East
Transporter	These data recorded names of transporters involved in cargo loss incidents, including various sea/air/land freight forwarders and express carriers	According to the characteristics of transport operators, transporter types were transformed into trucking company, express companies, and general international sea and air forwarders. Hence, a new category attribute (forwarder type) was generated
Claim payment	This field attribute recorded the monetary amount claimed from the insurance company, indicating the actual loss of cargo	A new category attribute (financial impact) was created to represent the severity of cargo loss based on practical management

Table IV. Data transformation

consuming and complex. In addition, a high compensation amount directly influences the insurance premium for the subsequent year, generating a strong economic impact, costs, and risks. Therefore, the case company viewed the level of severity for this type of loss as "high."

Predictive analytics

This study utilized SAS Enterprise Miner 12.1, a decision tree analysis module, to establish the classification model of cargo loss severity:

- The selection of target and input attributes: in this study, target attribute was set to be "financial impact," representing cargo loss severity. Moreover, input attributes were set to be "product category," "forwarder type," "on board date," "ship from," "ship to area," and "transit type."
- The separation of the cargo loss data set: the cargo loss data were allocated to three data sets used in this study, training data, validation data, and test data, at a proportion of 60, 30, and 10 percent, respectively.
- The establishment of the decision tree splitting rules: this study adopted the default splitting rules of SAS Enterprise Miner. The parameter settings can be seen in Table V.
- The development of the cargo loss classification model: this study developed a cargo loss classification model and obtained the analytical results of key decision tree indicators in Table VI. Furthermore, seven rules were summarized from the tree branches of the cargo loss severity classification model, as elaborated in Table VII.

Prescriptive analytics

To assess the robustness of findings, an identical cargo loss data set and variable settings were used at this stage, and another commonly used classification technique, logistic regression (Chen *et al.*, 2015), was employed for comparison. The logistic regression results

Table V.
Parameter settings in SAS Enterprise Miner 12.1

Parameter	Setting value	Description
Splitting criterion	Entropy	Splitting was performed using entropy reduction or other decision tree statistics. Obtained classification results were similar or consistent. Therefore, this study used entropy directly to perform splits
Maximal number of branches	3	Observation values were converged, predominantly yielding attribute values in ternary logic or higher. For example, transit type was divided into air, sea, and land. Excessive branching of a tree causes it to lose its managerial meaning. Therefore, the maximal number branches in this study was 3
Maximal depth of tree	6	The objective of this study was to provide practical references for use in transportation and logistics decision making. Restricting the maximal depth of a tree prevents a tree from overexpansion, which hinders interpretations and managerial decisions in practice. In addition, limited depth can avoid generating overly few training leaf nodes that would yield no substantial value and predictive power

Table VI.
The analytical results of key decision tree indicators

Indicator	Description
Total frequency	Training data: 782 pieces Validation data: 390 pieces Test data: 133 pieces
Classification error rate	Training data: 29.41% Validation data: 28.46% Test data: 27.81%
Variable importance	1. Transit type (1.0000) 2. Product category (0.455) 3. Ship to area (0.438)

indicated that the accuracy of the severity model was 69 percent, which was marginally lower than that generated using the decision tree method. Specifically, the severity results of the decision tree model exhibited an average accuracy rate of 71 percent (the accuracy rate of training data is 70.59 percent; that of validation data is 71.54 percent, and that of test data is 72.19 percent), demonstrating favorable classification accuracy. Furthermore, the evidence and comparison between the decision tree and logistic regression underpinned the validity and applicability of the decision tree analysis in classifying cargo loss issues.

The critical logistics factors affecting the severity of cargo loss are (in order of high to low importance) transit types, product categories, and shipping destinations. Overall, during decision tree splitting, transit type was determined to be the most crucial splitting attribute for the severity model, and a ternary tree for air, sea, and land transit types was generated. This classification result entailed extending differing transit types into independent branches, conforming to the norms of logistics management in which managerial decisions are made according to transit type. Moreover, product categories and shipping destinations also play an important role in determining the severity of cargo loss. Additionally, the remaining input data attributes (i.e. forwarder type, on board date, and ship from) within the data set yielded only tenuous discriminatory effects on classification operations; therefore, the decision tree model excluded the use of these input attributes.

The essential empirical results of the cargo loss business analytics are further discussed in "Risk management strategies for cargo loss prevention". Herein, managerial implications suggest what combination of logistics factors are more likely to trigger the severity of cargo loss. Since company resources are limited in practice, managers can pay close attention to the impact

Rule	Tree nodes	Rule descriptions	Accuracy rate of leaf nodes (%)	Coverage rate of observations (%)
S1	2	if transit type IS ONE OF: Air then Number of observations = 474 Predicted: financial impact = low = 0.70 Predicted: financial impact = medium = 0.27 Predicted: financial impact = high = 0.03	70	60.6
S2	8	if transit type IS ONE OF: Sea AND ship to area IS ONE OF: Asia, North America, Africa then Number of observations = 77 Predicted: financial impact = low = 0.82 Predicted: financial impact = medium = 0.18 Predicted: financial impact = high = 0.00	82	9.8
S3	9	if transit type IS ONE OF: sea AND Ship to area IS ONE OF: Europe then Number of observations = 68 Predicted: financial impact = low = 0.59 Predicted: financial impact = medium = 0.29 Predicted: financial impact = high = 0.12	59	8.7
S4	10	if transit type IS ONE OF: sea AND Ship to area IS ONE OF: Oceania, Middle-East then Number of observations = 21 Predicted: financial impact = low = 0.29 Predicted: financial impact = medium = 0.57 Predicted: financial impact = high = 0.14	57	2.7
S5	11	if transit type IS ONE OF: Truck AND product category IS ONE OF: NB then Number of observations = 98 Predicted: financial impact = low = 0.17 Predicted: financial impact = medium = 0.83 Predicted: financial impact = high = 0.00	83	12.5
S6	12	if transit type IS ONE OF: truck AND product category IS ONE OF: MB VGA then Number of observations = 29 Predicted: financial impact = low = 0.34 Predicted: financial impact = medium = 0.48 Predicted: financial impact = high = 0.17	48	3.7
S7	13	if transit type IS ONE OF: truck AND product category IS ONE OF: desktop, LCD monitor, Others then Number of observations = 15 Predicted: financial impact = Low = 0.67 Predicted: financial impact = medium = 0.33 Predicted: financial impact = high = 0.00	67	1.9

Notes: Coverage rate was calculated by dividing the number of node training data by 780 pieces of root node original training data. For example, Rule S1 involves 474 pieces of leaf node training data, with a coverage rate of 474/780 = 60.6 percent

Table VII.
Essential rules for cargo loss severity

of logistics operations on the severity of cargo loss. Moreover, managers can analyze their cargo loss incidents by the proposed cargo loss business analytics, and prioritize management resources to avert cargo loss risks.

Risk management strategies for cargo loss prevention

Based on the empirical results of the cargo loss business analytics, this study devised strategies for cargo loss prevention and managerial recommendations for executing cargo loss mitigation plans.

Deter cargo theft when shipping high-value products via land transportation

Land transit of high-value products is more likely to provoke cargo theft based on the experience of the case company. Rule S5 (Table VII) indicates that when laptops are shipped via land transportation modes, cargo loss of medium severity is most likely to occur, exhibiting an accuracy rate of 83 percent. Moreover, Rule S6 (Table VII) shows that when electronic card products are shipped on land, cargo loss of medium severity is most likely to occur, with an accuracy rate of 48 percent. However, noticeably, the probability of highly severe cargo loss occurring under Rule S6 was 17 percent, the highest percentage among all lead nodes generated in this loss severity model. Rule S6 (land transit involving electronic card products) exhibited a possibility of a high degree of loss greater than that under Rule S5 (land transport shipping of laptop products).

The above findings provide extremely valuable information to managers: transporting laptops and electronic card products on land is more likely to precipitate severe losses due to cargo thefts. Electronic products such as laptops are small in size, expensive, and highly mobile. Hence, logistics managers should strictly regulate their control of land forwarders. Furthermore, formulating rigorous standard logistics procedures and penalty strategies are vital loss prevention approaches to avoid leaking information about shipping high-value products. Besides, since cargo thieves may change their modus operandi occasionally, logistics managers should update their cargo loss mitigation strategies frequently.

Cautiously utilize ocean shipment to certain areas

Even though shipping through sea transport is the cheapest way for transnational logistics, it might not have a cost advantage once cargo loss ramifications are taken into account. According to Rule S4 (Table VII), when products are shipped using sea transport to Australia or the Middle-East, cargo loss of medium severity is likely to occur, with an accuracy rate of 57 percent. Besides, examining the decision tree rules from a probability perspective reveals that Rule S4 resulted in 14 percent highly severe cargo loss incidents, which is substantially higher than that of the root node (4 percent). This result indicates that extremely severe financial loss is highly likely to occur under such a logistics condition, and therefore logistics managers must strengthen control for this condition as outlined in rule S4.

Since not all areas are very appropriate for maritime logistics, managers making logistics decisions should pay close attention to cargo loss risks when shipments are made via sea transport to the Australian or Middle-East regions. For example, maritime logistics may strengthen cargo packaging and avoid fictitious pickups at collection ports.

Allocate resources to critical logistics factors

This study found that forwarder type was not selected during the process of tree splitting as part of a category in the decision tree. Specifically, no significant difference was observed in cargo loss severity based on forwarder type (air and sea forwarders, express companies, or trucking companies). This finding implies that the freight forwarders have similar abilities to alleviate cargo loss, especially for the case company. Accordingly, managers can scrutinize their cargo loss data and then allocate resources from selecting the best forwarder to other crucial factors engendering cargo loss (e.g. transit type, product category, and shipping destination) when implementing cargo loss prevention.

Devise strategic cargo loss insurance policies

Cargo loss insurance policies could be developed differently based on a compound of logistics activities. Most insurance policies usually treat the insured according to the accumulated amount of insurance claims. Nevertheless, the empirical results of this study reveal that cargo loss risk can be further identified by different cargo loss rules. Consequently, how to set up appropriate insurance policies turns out to be important. Managers can propose strategic insurance policies to insurance firms based on their own cargo loss data. Specifically, utilizing the proposed business analytics of cargo loss can help companies to identify their risk chains (various levels of cargo loss severity) with different combinations of logistics risk ingredients. Additionally, stakeholders involved in a risk chain could share risk (liability) and arrange a grouping of cargo loss insurance.

Conclusions

Cargo loss incidents not only cause financial losses to an enterprise but also disrupt the whole logistics systems. This study gives academics and managers a holistic view of cargo loss in global logistics systems through the data-driven analytics of a case company. Furthermore, logistics companies can utilize the proposed business analytics framework to identify cargo loss issues and thereby to develop vital cargo loss prevention strategies for risk management of logistics systems.

This study differs from previous investigations by addressing the logistics risk management in numerous ways. First, this pioneer empirical study substantially contributes to management of cargo loss with the framework of business analytics involving descriptive, predictive, and prescriptive analyses. Second, this study systematically absorbs the valuable insights of both academics and practitioners on the issues of cargo loss analysis and prevention. Third, this study proposes a business analytics framework by integrating the KDD procedure with big data analytics to tackle cargo loss issues in logistics systems. Fourth, the empirical results of the case study reveal that transit types, product categories, and shipping destinations are key factors behind cargo loss severity. Finally, this study devised strategies for risk management in logistics systems as follows: “deter cargo theft when shipping high-value products via land transportation,” “cautiously utilize ocean shipment to certain areas,” “allocate resources to critical logistics factors,” and “devise strategic cargo loss insurance policies.” This pioneer empirical study explores valuable cargo loss data from the insurance reimbursement claim database of the case company’s insurance company to help not only prevent financial losses incurred by cargo loss incidents but also avoid jeopardizing an enterprise’s competitiveness.

Future studies could help determine whether physical characteristics of cargo products (e.g. packaging methods, strength of packaging materials, shipment volume, and weight, etc.) are more or less associated with cargo loss incidents. The current findings may be limited to companies with characteristics similar to those of the case company. Various real cargo loss cases can be further investigated to create general risk management strategies for cargo loss prevention. Moreover, a compromise solution of an insurance contract among cargo owners, logistics service providers, and insurance companies could be scrutinized in future studies.

References

Archak, N., Ghose, A. and Ipeirotis, P.G. (2011), "Deriving the pricing power of product features by mining consumer reviews", *Management Science*, Vol. 57 No. 8, pp. 1485-1509.

Burges, D. (2012), *Cargo Theft, Loss Prevention, and Supply Chain Security*, Butterworth-Heinemann Ltd, Oxford.

Cameron, V. and Sheridan, J. (2013), "Stock throughput policies", *An Industry Seminar: Marine Insurance Day*, American Institute of Marine Underwriters, New York, NY.

Chen, M.C., Chang, K.C., Hsu, C.L. and Xiao, J.H. (2015), "Applying a Kansei engineering-based logistics service design approach to developing international express services", *International Journal of Physical Distribution & Logistics Management*, Vol. 45 No. 6, pp. 618-646.

Chopra, S. and Sodhi, M.S. (2014), "Reducing the risk of supply chain disruptions", *MIT Sloan Management Review*, Vol. 55 No. 3, pp. 73-80.

Davis, C.K. (2014), "Beyond data and analysis", *Communications of the ACM*, Vol. 57 No. 6, pp. 39-41.

Durach, C.F., Wieland, A. and Machuca, J.a.D. (2015), "Antecedents and dimensions of supply chain robustness: a systematic literature review", *International Journal of Physical Distribution & Logistics Management*, Vol. 45 Nos 1-2, pp. 118-137.

Ekwall, D. (2009), "The displacement effect in cargo theft", *International Journal of Physical Distribution & Logistics Management*, Vol. 39 No. 1, pp. 47-62.

Ekwall, D. and Lantz, B. (2013), "Seasonality of cargo theft at transport chain locations", *International Journal of Physical Distribution & Logistics Management*, Vol. 43 No. 9, pp. 728-746.

Evans, J.R. (2012), "Business analytics: the next frontier for decision sciences", *Decision Line*, Vol. 43 No. 2, pp. 4-6.

Fayyad, U. and Stolorz, P. (1997), "Data mining and KDD: promise and challenges", *Future Generation Computer Systems*, Vol. 13 Nos 2-3, pp. 99-115.

Fennelly, L. (2012), *Handbook of Loss Prevention and Crime Prevention*, Butterworth-Heinemann, Oxford.

Figueiredo, V., Rodrigues, F., Vale, Z. and Gouveia, J.B. (2005), "An electric energy consumer characterization framework based on data mining techniques", *IEEE Transactions on Power Systems*, Vol. 20 No. 2, pp. 596-602.

FreightWatch (2013), *2013 Global Cargo Theft Threat Assessment*, FreightWatch International, Austin, TX.

Hale, T. and Moberg, C.R. (2005), "Improving supply chain disaster preparedness: a decision process for secure site location", *International Journal of Physical Distribution & Logistics Management*, Vol. 35 No. 3, pp. 195-207.

Hohenstein, N.O., Feisel, E., Hartmann, E. and Giunipero, L. (2015), "Research on the phenomenon of supply chain resilience: a systematic review and paths for further investigation", *International Journal of Physical Distribution & Logistics Management*, Vol. 45 Nos 1-2, pp. 90-117.

Kern, D., Moser, R., Hartmann, E. and Moder, M. (2012), "Supply risk management: model development and empirical analysis", *International Journal of Physical Distribution & Logistics Management*, Vol. 42 No. 1, pp. 60-82.

Kokotos, D.X. and Smirlis, Y.G. (2005), "A classification tree application to predict total ship loss", *Journal of Transportation and Statistics*, Vol. 8 No. 2, pp. 31-42.

Kutz, M. (2007), *Handbook of Farm, Dairy, and Food Machinery*, William Andrew Publishing, New York, NY.

Lai, L.H. (2008), "An evaluation of fuzzy transportation underwriting systematic risk", *Transportation Research Part A: Policy and Practice*, Vol. 42 No. 9, pp. 1231-1237.

Lambert, D.M., Stock, J.R. and Ellram, L.M. (1998), *Fundamentals of Logistics Management*, McGraw-Hill Publishing, Singapore.

Manuj, I. and Mentzer, J.T. (2008), "Global supply chain risk management strategies", *International Journal of Physical Distribution & Logistics Management*, Vol. 38 Nos 3-4, pp. 192-223.

Mookherjee, R., Mukherjee, J., Martineau, J., Xu, L., Gullo, M., Zhou, K., Hazlewood, A., Zhang, X., Griarte, F. and Li, N. (2016), "End-to-end predictive analytics and optimization in Ingram Micro's two-tier distribution business", *Interfaces*, Vol. 46 No. 1, pp. 49-73.

Phillips-Wren, G. and Hoskisson, A. (2015), "An analytical journey towards big data", *Journal of Decision Systems*, Vol. 24 No. 1, pp. 87-102.

Reich, D., Winkler, S.L., Klampfl, E. and Olson, N. (2015), "Ford uses analytics to help fleet customers buy more sustainable vehicles", *Interfaces*, Vol. 45 No. 6, pp. 543-553.

Roiger, R.J. and Geatz, M.W. (2003), *Data Mining: A Tutorial-Based Primer*, Addison-Wesley, Boston, MA.

Schoenherr, T. and Speier-Pero, C. (2015), "Data science, predictive analytics, and big data in supply chain management: current state and future potential", *Journal of Business Logistics*, Vol. 36 No. 1, pp. 120-132.

Skorna, A.C.H. and Fleisch, E. (2012), *Loss Prevention in Transportation to Ensure Product Quality: Insights from the Cargo Insurance Sector*, in Frick, J. and Laugen, B.T. (Eds), *Advances in Production Management Systems. Value Networks: Innovation, Technologies, and Management*, APMS 2011, IFIP Advances in Information and Communication Technology, Vol. 384, Springer, Berlin, Heidelberg.

Souza, G.C. (2014), "Supply chain analytics", *Business Horizons*, Vol. 57 No. 5, pp. 595-605.

Transported Asset Protection Association (2012), "New air cargo security standard takes TAPA theft prevention measures further up the supply chain", The Monthly Cargo Crime Update for Members of TAPA EMEA, FL.

UK P&I Club (2013), "Cargo damage – the causes", available at: www.ukpandi.com/fileadmin/uploads/uk-pi/LP%20Documents/Carefully_to_Carry/Cargo%20Damage.pdf (accessed May 9, 2016).

Venkatachalam, S., Wong, F., Uyar, E., Ward, S. and Aggarwal, A. (2015), "Media company uses analytics to schedule radio advertisement spots", *Interfaces*, Vol. 45 No. 6, pp. 485-500.

Vilko, J.P.P. and Hallikas, J.M. (2012), "Risk assessment in multimodal supply chains", *International Journal of Production Economics*, Vol. 140 No. 2, pp. 586-595.

Wieland, A. and Wallenburg, C.M. (2013), "The influence of relational competencies on supply chain resilience: a relational view", *International Journal of Physical Distribution & Logistics Management*, Vol. 43 No. 4, pp. 300-320.

Zanakis, S.H. and Becerra-Fernandez, I. (2005), "Competitiveness of nations: a knowledge discovery examination", *European Journal of Operational Research*, Vol. 166 No. 1, pp. 185-211.

About the authors

Pei-Ju Wu is an Assistant Professor in the Department of Transportation Technology and Management as well as the Chief Innovation & Development Officer of Innovation Center for Intelligent Transportation and Logistics at the Feng Chia University, Taichung, Taiwan. His research interests are in the area of logistics and supply chain management, operations management, and business analytics. Pei-Ju Wu is the corresponding author and can be contacted at: wupj@fcu.edu.tw

Mu-Chen Chen is a Professor in the Department of Transportation and Logistics Management as well as a Professor of the Degree Program of Transportation and Logistics at the College of Management, National Chiao Tung University, Taipei, Taiwan. His research interests are in the area of logistics and supply chain management, big data analytics and data mining, electronic commerce, service quality management, and meta-heuristics.

Chih-Kai Tsau received the MS Degree in Transportation and Logistics from the National Chiao Tung University, Taipei, Taiwan, in 2014. His research interests are in the area of global logistics management, procurement and materials management, and supply chain management.