# Data-driven fraud detection in international shipping

Ron Triepels [a,*], Hennie Daniels [a,b], Ad Feelders [c]

[a] *Tilburg University, Warandelaan 2, Tilburg, The Netherlands*
[b] *Erasmus University, Burgemeester Oudlaan 50, Rotterdam, The Netherlands*
[c] *Utrecht University, Princetonplein 5, Utrecht, The Netherlands*

## ARTICLE INFO

## ABSTRACT

Document fraud constitutes a growing problem in international shipping. Shipping documentation may be deliberately manipulated to avoid shipping restrictions or customs duties. Well-known examples of such fraud are miscoding and smuggling. These are cases in which the documentation of a shipment does not correctly or entirely describe the goods in transit. In an attempt to reduce the risks of document fraud, shipping companies and customs authorities typically perform random audits to check the accompanying documentation of shipments. Although these audits detect many fraud schemes, they are quite labor intensive and do not scale to the massive amounts of cargo that is shipped each day. This paper investigates whether intelligent fraud detection systems can improve the detection of miscoding and smuggling by analyzing large sets of historical shipment data. We develop a Bayesian network that predicts the presence of goods on the cargo list of shipments. The predictions of the Bayesian network are compared with the accompanying documentation of a shipment to determine whether document fraud is perpetrated. We also show how a set of discriminative models can be derived from the topology of the Bayesian network and perform the same fraud detection task. Our experimental results show that intelligent fraud detection systems can considerably improve the detection of miscoding and smuggling compared to random audits.

## 1. Introduction

Trade liberalization and technological innovation have considerably changed the international shipping industry over the last century. Nowadays, on average 350 thousand TEU's[1] of containerized cargo is shipped across the world each day (World Shipping Council, 2004). Such excessive demand is detrimental to shipping companies and customs authorities to guarantee safe and compliant operations. Shipping companies often need to process shipments without knowing the exact nature of the goods inside a box or container (Hesketh, 2010), while customs authorities can only physically inspect a fraction of the shipments that cross the borders of a country. This leaves room for fraudsters to perpetuate all kinds of fraudulent activities.

Fraud in international shipping occurs in many forms and on different scales, ranging from local cargo theft to international smuggling. Either way, tracks of a fraud scheme must be covered in the documentation of a shipment. This form of fraud is also known as document fraud. Document fraud is the act of manipulating facts in contracts or agreements with the intent to benefit by commercial gain (Hill & Hill, 2009). The most common types of document fraud in international shipping are miscoding and smuggling.

Miscoding refers to the act of providing incorrect information about goods in transit. Knowing the exact nature of goods that cross the borders of a country is essential for customs authorities, as this information constitutes the basis for enforcing shipping restrictions and levying customs duties. Therefore, contracting parties in a shipment are obliged to classify goods in transit according to an internationally accepted coding scheme called the Harmonized System (HS).[2] Based upon this classification, customs agents decide under which conditions goods are allowed to be transported across countries and how much customs duties the importer or exporter needs to pay. Miscoding occurs when a party specifies

* Corresponding author.
*E-mail addresses:* r.j.m.a.triepels@uvt.nl (R. Triepels), h.a.m.daniels@uvt.nl (H. Daniels), a.j.feelders@uu.nl (A. Feelders).

[1] Twenty-foot Equivalent Unit (TEU) is a measure used within the international shipping industry to denote the capacity of a cargo container. One TEU equals a 20-foot-long intermodal container.

[2] The Harmonized System is an international product nomenclature introduced by the World Customs Organization. It captures about five thousand commodity groups which are identified by six-digit codes.

HS-codes of other goods with similar properties but which are not prohibited or require to pay lower customs duties.

In contrast, smuggling refers to the act of secretly shipping goods under conditions that are against the law by any country that is crossed by the shipment. Smuggled goods are usually put inside a shipment somewhere along the supply chain while making sure that they are not listed on any official documentation provided to local customs authorities. Once the shipment has been cleared in the destination country, the smuggled goods are secretly removed from the shipment to avoid any customs regulations. Drugs, weapons, cigarettes, and alcohol are examples of goods that are frequently smuggled because they are prohibited or require to pay higher amounts of customs duties.

To mitigate the risks of document fraud, shipping companies and customs authorities perform random audits to check the accompanying documentation of shipments. For example, shipping companies have experienced customs officers that check whether the bills of lading and trade certificates issued for a shipment are valid and consistent. Also, customs authorities perform physical inspections and x-ray scans at customs borders to check whether a box or container contains those goods listed on the corresponding customs declaration. Although many fraud schemes are detected by such audits, they do not scale well to the vast amount of cargo that is processed on a daily basis.

It is believed that intelligent systems can significantly improve the detection of fraud in international supply chains (Gordhan, 2007). Intelligent systems are systems that emulate the decision-making ability of human experts by analyzing large sets of data using statistical techniques and, more recently, machine learning techniques (Aronson, Liang, & Turban, 2005). Instead of choosing shipments randomly, intelligent systems can be employed to analyze the vast amount of data that is generated by supply chains and select only potential fraudulent shipments for further fraud analysis. In this way, supply chain participants can better allocate their limited resources for fraud detection. Several systems have been proposed for this purpose. However, it is unclear to which extend such systems do indeed improve the detection of document fraud.

In this paper, we investigate the extend to which intelligent fraud detection systems can improve the detection of miscoding and smuggling compared to random audits. We first develop a Bayesian network that detects miscoding and smuggling by analyzing trade patterns and itinerary patterns in shipment data. Bayesian networks are probabilistic generative models that have been successfully applied in many fraud detection tasks, see e.g. Ezawa and Schuermann (1995), Taniguchi, Haft, Hollmén, and Tresp (1998) and Kirkos, Spathis, and Manolopoulos (2007). Accordingly, we discuss how different probabilistic discriminative models can be derived from the topology of the Bayesian network. We evaluate the performance of the models and compare their predictions with a set of random audits that generate the same amount of alarms. Our results confirm that intelligent fraud detection systems can select shipments for further fraud analysis much better than random audits.

## 2. Related research

In this section, we provide a brief overview of related research on the detection of document fraud in international shipping. We discuss how document fraud is detected by analyzing trade patterns (Section 2.1) and itinerary patterns (Section 2.2). Furthermore, we introduce a hybrid approach based on the analysis of both types of patterns (Section 2.3) and compare its main features with existing fraud detection models in the literature (Section 2.4).

### 2.1. Trade-based fraud detection

One way to detect document fraud is to analyze deviations in the cargo that is traded between importers and exporters. We will refer to this approach as trade-based fraud detection. The objective of trade-based fraud detection is to find deviating trade patterns, i.e. cases were countries or organizations engage in trade that deviates from the type of goods that are usually traded, or involves goods with extraordinary properties like their price or weight.

Several models have been proposed to detect deviating trade patterns in historic customs declarations. Filho and Wainer (2007) built a hierarchical Bayesian classifier to predict document fraud. The main idea behind their classification model is to model combinations of binary features, e.g., an HS-code and country of origin, in a hierarchical structure such that there is strong independence between the feature combinations while the most specific ones dominate the classification. Yaqin and Yuming (2010) built a classification model based on association rule mining. Association rule mining is performed separately on the set of fraudulent and non-fraudulent declarations while keeping the class as the antecedent. Their model classifies declarations by determining the class of the association rule that matches the declaration and has the highest confidence and support. Digiampietri et al. (2008) proposed a visual anomaly detection system to detect document fraud. Their system compares features of declared goods with features of similar goods declared by the importer in the past. When similar goods are found, combinations of features, e.g., price and weight, are retrieved and highlighted in diagrams. These diagrams need to be visually inspected to determine the extent to which goods deviate from the expected norm. Finally, Hua, Li, and Tao (2006) proposed a classification model based on clustering and logistic regression. Their model groups declarations into approximately homogeneous clusters based on the prices and weights of the goods declared. Accordingly, for each cluster, a logistic regression function is fitted that predicts document fraud based on a set of highly correlated features.

### 2.2. Itinerary-based fraud detection

Another way to detect document fraud is to analyze deviations in the way that cargo is shipped through the global shipping network. We will refer to this approach as itinerary-based fraud detection. The objective of itinerary-based fraud detection is to find deviating itinerary patterns, i.e. cases where goods are shipped via itineraries that are not very economically beneficial. Such patterns are often found by analyzing digital shipping messages. Shipping messages are created by shippers and shared across a shipping network to inform others about the status and movement of a shipment. These messages typically include details about the location of a shipment at a given moment in time and its status, e.g., arrival or transshipment.[3]

Several studies have investigated how we can find deviating itinerary patterns in shipping messages. Chahuara et al. (2014) address the problem of the heterogeneous nature of container events and its negative impact on the analysis of itineraries. Shipping messages are collected from various sources and can be ambiguous, incomplete, imprecise or redundant. To deal with this noise, the researchers built a conditional random field to classify the status of container messages based on a set of spatiotemporal features. Villa and Camossi (2011) built an ontology of the maritime container domain. Their ontology defines objects such as a

---

[3] Transhipment is the process of shipping goods to an intermediate location from which they proceed their journey. Usually, transshipment is performed to change between vessels with pre-defined routes or to change the mode of transport.