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*

A B S T R A C T

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.
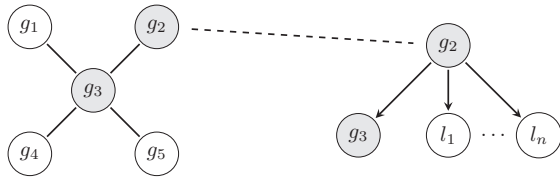
Fig. 1. A schematic overview of the model proposed in Triepels et al. (2015). The left graph is a Markov random field on a set of goods. The right graph is a Bayesian network classifier that is constructed for good g_2 having the Markov blanket of the good in the Markov random field and a set of itinerary (location) variables as explanatory variables.

container or vessel, processes such as import or export, and relationships between these objects and processes. These semantics are applied in combination with logical predicates to perform reasoning on anomalous itinerary patterns. Camossi, Dimitrova, and Tsois (2012) showed how to detect deviating itinerary patterns by a support vector machine in one-class classification. Their model uses spatiotemporal features to model the way shipments normally find their way through the global shipping network. Deviating itineraries are identified by determining the extent to which recent itineraries deviate from the expected norm. Finally, Dimitrova, Tsois, and Camossi (2014) developed a web-based system to visualize global shipping traffic. Their system retrieves shipping messages from historically taken itineraries and plots the coordinates of locations that are crossed in the itineraries on a geographical map. Filter and aggregation functions can be applied to study frequently reoccurring patterns in the itineraries.

2.3. Hybrid approach

Trade patterns and itinerary patterns constitute a valuable source of information to detect potential cases of document fraud. However, fraud detection models usually analyze these patterns separately, while they are closely related to each other in many fraud cases. For example, a common fraud practice is to use transshipment to conceal the origin of cargo (World Customs Organization, 2012). Such fraud is difficult to detect. Trade patterns alone will not highlight the deviance in the itinerary, while itinerary patterns alone will highlight the deviance but lack the information to determine whether the transshipment is justifiable[4] For this reason, combining both types of patterns poses an important challenge for fraud detection models in international shipping.

In Triepels, Feelders, and Daniels (2015), we proposed a model that detects miscoding and smuggling by simultaneously analyzing trade patterns and itinerary patterns in shipment data. Fig. 1 provides an overview of the model construction. The model is constructed in two steps. First, feature selection is performed to determine the types of goods which are statistically independent of the presence of other goods in a shipment. These independencies are identified by constructing a Markov random field on a set of binary variables indicating the presence of a particular good in a shipment, and for each good, determine its Markov blanket in the Markov random field. Accordingly, a Bayesian network classifier is constructed for each good that predicts the presence of the good in a shipment based on the presence of other dependent goods and a set of itinerary (location) variables. Finally, miscoding and smuggling incidents are detected by determining whether there is a mismatch between the goods predicted by the classifiers and the goods listed on the cargo lists of shipments.

This paper provides an alternative solution to address the same problem. We model shipments directly in a Bayesian network by a set of variables representing both cargo details as well as itinerary details. The advantage of this approach is that it makes the separate feature selection step redundant. Instead, we apply the Bayesian network to automatically perform feature selection for all goods at once. Moreover, we can derive discriminative classifiers for each good from the topology of the network and use these to perform the fraud detection task instead. Experimental tests reveal that these discriminative classifiers tend to generate alarms for miscoding and smuggling with higher precision and recall compared to the Bayesian network.

2.4. Model features

The main innovation of our Bayesian network is that it analyzes both trade patterns and itinerary patterns to detect document fraud. Besides this feature, the network offers several additional features that are relevant to the international shipping industry. First, the network can predict miscoding and smuggling based on incomplete shipment data and update its predictions when additional information about a supply chain becomes available. Many predictive models do not support this feature. Second, the network allows expert knowledge of customs agents to be incorporated into the fraud predictions. Domain experts can help constructing the conditional independence structure of a Bayesian network (Cowell, Dawid, Lauritzen, & Spiegelhalter, 2006) or provide information about the quantitative influences of the nodes to improve the model estimation (Feelders, 2012). Third, the conditional independence structure of the network can visually aid the tasks of customs agents. Finally, the network can be applied to perform probabilistic reasoning on shipment behavior and generate fraud alarms that are easy to interpret. Table 1 summarizes these features and shows the extent to which they are supported by existing models in the literature.

3. Detection of document fraud by Bayesian networks

In this section, we elaborate on the details of our fraud detection model. We introduce some concepts of international shipping (Section 3.1) and formalize the detection problem of miscoding and smuggling (Sections 3.2 and 3.3). Furthermore, we define a Bayesian network and discuss how it can be applied to detect these forms of fraud (Sections 3.4 and 3.5). Finally, we show how a set probabilistic discriminative models can be derived from the topology of the Bayesian network and applied to perform the same fraud detection task (Section 3.6).

3.1. Shipping concepts

Let $\mathcal{G}$ be the set of all internationally standardized commodity codes[5] Moreover, let $\mathcal{L}$ be the set of all locations between which goods are transported. An important shipping concept is the cargo list.

Definition 1. A cargo list $C = \{g_1, \ldots, g_k\}$ is a subset of $\mathcal{G}$, where each $g_i \in C$ is a good with commodity code i that is conveyed by a shipment, and k is an integer denoting the size of the cargo list.

Goods on the cargo list are transported via an itinerary through the global shipping network. The itinerary usually involves multiple shipping companies that each takes care of a specific part of the itinerary, possibly by a different mode of transportation. We define an itinerary as a set of locations that are crossed by a shipment in the global shipping network.

---

[4] Transshipment may, for example, be more likely for small and cheap goods (e.g. phone accessories) than for large and expensive goods (e.g. furniture).

[5] Currently, only the first six-digits of the HS nomenclature are internationally standardized.

Table 1
A comparison of features supported by the model proposed in this paper and existing models in the literature.

Model	Features					
	Analyzes:		Supports:		Provides:	
	Trade patterns	*Itinerary patterns*	*Missing data*	*Expert knowledge*	*Fraud probabilities*	*Visual aid*
Filho and Wainer (2007)	Yes	No	Yes	No	Yes	No
Yaqin and Yuming (2010)	Yes	No	Yes	No	No	No
Digiampietri et al. (2008)	Yes	No	No	No	No	Yes
Hua et al. (2006)	Yes	No	No	Yes	Yes	No
Chahuara et al. (2014)	No	Yes	Yes	No	No	No
Villa and Camossi (2011)	No	Yes	Yes	Yes	No	No
Camossi et al. (2012)	No	Yes	No	No	No	Yes
Dimitrova et al. (2014)	No	Yes	Yes	Yes	No	Yes
Proposed model	Yes	Yes	Yes	Yes	Yes	Yes

Country of Origin   Destination Country
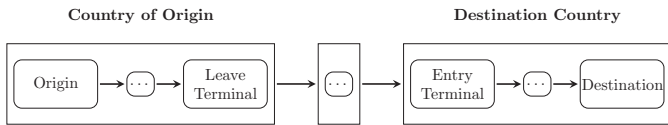
Fig. 2. The general structure of an itinerary in international shipping.

Definition 2. An itinerary $I = <l_1, \ldots, l_n>$ is a ordered subset of $\mathcal{L}$, where each $l_i \in \mathcal{L}$ represents a location that is crossed by a shipment, and n denotes the length of the itinerary. We denote the set of all itineraries by $\mathcal{I}$.

The order in which locations in the itinerary are crossed matters, so $<l_1, l_2> \neq <l_2, l_1>$. Furthermore, itineraries may be of variable length. They usually consist of a sequence of locations corresponding to shipping terminals that are crossed during the journey, like ports, airports, truck terminals, or railway stations.

Itineraries in international shipping have a general structure as shown in Fig. 2. They consist of at least three smaller transportation parts. First, cargo is picked up at the origin and distributed within the country of origin by in-land transportation. Accordingly, the cargo is moved to the destination country by cross-border transportation. This part is typically performed by sea or air transport and consists of multiple transports that move the cargo across intermediate countries. Finally, when the cargo reaches its destination country, it is distributed to its final destination by inland transportation. Some locations in the itinerary have a special interpretation. Usually, the first and last location represent the origin and destination, while the locations connecting the in-land and cross-border transportation are respectively the leave and entry terminals.

A shipment consists of a specific cargo list and itinerary, along with an indication of the shipment duration. It reflects the conditions under which goods are transported from the origin to the destination.

Definition 3. A shipment $s = (C_s, I_s, T_s)$ is a triple, where $C_s \in \mathcal{G}$ denotes the cargo list of the shipment, $I_s \in \mathcal{I}$ the itinerary of the shipment, and T_s the shipment duration. We denote the set of all shipments by $\mathcal{S}$.

3.2. Fraud detection task

Miscoding and smuggling can be detected by looking at the probability of goods being listed on the cargo list of a shipment. If it is improbable that a shipment conveys a good on the cargo list, then it might be subject to miscoding. Similarly, if it is probable that a good is conveyed by a shipment but missing on the cargo list, then it might be subject to smuggling.

We denote the probability of goods being conveyed by a shipment by function $\mathbf{P}$:

$$\mathbf{P} : \mathcal{S} \to [0, 1]^{|\mathcal{G}|} \qquad (1)$$

$\mathbf{P}(s)$ is a vector of probabilities where each $\mathbf{P}(s)_i$ denotes the probability that good i is conveyed by s. Furthermore, let $\psi_1 : \mathcal{S} \to \{0, 1\}$ and $\psi_2 : \mathcal{S} \to \{0, 1\}$ be two functions that assign a binary value to a shipment that indicates whether they are likely subject to miscoding and smuggling respectively. Shipments are classified as potentially fraudulent if the goods on their cargo list deviate from the expected goods estimated by $\mathbf{P}$. Function ψ_1 classifies s as being subject to miscoding when it contains a good i on the cargo list for which $\mathbf{P}(s)_i$ is low:

$$\psi_1(s) = \begin{cases} 1 & \text{if } \exists g_i \in C_s (\mathbf{P}(s)_i \leq \alpha) \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

Here, $\alpha \in (0, 1)$ is a risk threshold close to zero. Likewise, function ψ_2 classifies s as being subject to smuggling when there exists a good that is not on the cargo list but for which $\mathbf{P}(s)_i$ is high:

$$\psi_2(s) = \begin{cases} 1 & \text{if } \exists g_i \notin C_s (\mathbf{P}(s)_i \geq \beta) \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

Here, $\beta \in (0, 1)$ is a risk threshold close to one. α and β determine the confidence level at which ψ_1 and ψ_2 respectively infer that fraud is perpetuated. They can be adjusted to meet the level of risk tolerance.

3.3. Estimation of $\mathbf{P}$

We want to estimate the probability that a shipment conveys specific types of goods given its cargo list, itinerary, and shipment duration. For an individual good g_i, this probability can be defined as a conditional probability:

$$\mathbf{P}(s)_i = P(g_i|s) = P(g_i|C_s \setminus g_i, I_s, T_s) \qquad (4)$$

where, $\mathbf{P}(s)_i$ is the probability that g_i is present in C_s given all other goods $C_s \setminus g_i$ on the cargo list, the locations in the itinerary I_s, and the shipment duration T_s. We estimate $\mathbf{P}$ from a dataset of historical shipments $D \subset \mathcal{S}$, under the assumption that the majority of the shipments in D are correctly declared and are not fraudulent.

Estimating $\mathbf{P}$ directly from D is problematic. This problem arises from the fact that shipment data is high dimensional. It compromises a vast amount of ways in which goods can be shipped from one location to another. To quantify the probability of goods being conveyed by a shipment, we would need an enormous sample to observe the exact shipment multiple times and count the goods listed on its cargo list. In practice, however, the number of shipments in data is usually not large enough to get reliable estimates by a simple counting approach. To avoid this problem, we model shipments in a Bayesian network and apply inference to deduce the conditional probabilities of $\mathbf{P}$.

3.4. Bayesian networks[6]

A Bayesian network (BN) over a set of random variables $X = \{x_1, \ldots, x_m\}$ can be defined as a tuple $\mathcal{BN} = (\mathcal{N}, \Theta)$ where, $\mathcal{N} = (V, E)$ is a directed acyclic graph whose nodes V index X and edges E represent dependencies among the variables, and Θ is a set of parameters such that $\theta_v \in \Theta$ defines the conditional probability of x_v given its parents in $\mathcal{N}$ (Koller & Friedman, 2009). A nice property of a BN is that it allows to estimate the joint probability distribution $P(X)$ efficiently. Instead of estimating the probability of each possible configuration of the variables in X, a BN assumes conditional independence structure $\mathcal{N}$ on X and estimates $P(X)$ as the product of each x_v conditioned on its parents:

$$P(X) = \prod_{v \in V} P(x_v | \mathrm{PA}(x_v)) \qquad (5)$$

where $\mathrm{PA}(x_v)$ is the set of parents of x_v in $\mathcal{N}$. When $\mathcal{N}$ is a good representation of the independence structure of X, a BN can provide a better estimate of $P(X)$ through the estimation of less and more reliable (conditional) probabilities.

The Markov blanket plays an important role to understand independence structure $\mathcal{N}$. The Markov blanket of x_v, denoted as $MB(x_v)$, is the set of x_v's parents, its children, and the parents of its children (Pearl, 1988). Because a BN estimates $P(X)$ by the factorization in Eq. (5), it can be shown that each x_v is conditionally independent of the rest of the variables in the network given $MB(x_v)$. In other words, $MB(x_v)$ defines the boundary that shields x_v from the probabilistic influence of the remaining variables in the network.

3.5. A Bayesian network of shipments

There are multiple ways to represent shipments in a BN. We discuss two possible options which we call the Mixed Shipment Network (MSN) and Binary Shipment Network (BSN). In the MSN, shipment are represented by a combination of binary and multinomial variables. This includes a binary variable g_i for each good denoting whether it is present on the cargo list, a multinomial variable l_j for each jth position of the itinerary denoting the location crossed at this particular position in the itinerary,[7] and a multinomial variable t_s denoting the shipment duration. The BSN is similar to the MSN except that it also represents the itinerary and shipment duration by a set of binary variables. This includes a binary variable g_i for each good, a binary variable l_j^i for each location i at the jth position of the itinerary,[8] and a binary variable t_i for each shipment duration. Figs. 3 and 4 highlight these differences. They depict the Markov blanket corresponding the same good in respectively an MSN and BSN constructed from real-world shipment data.

Both types of networks model the same information about a shipment but may provide a different estimate of **P** due to the different granularity at which they model conditional independencies. Because of the independence property of the Markov blanket, we can estimate the presence of each good on the cargo list as:

$$\mathbf{P}(s)_i = P(g_i | MB(g_i)) \qquad (6)$$

The BSN captures conditional independencies at the instance level (between individual goods, itinerary locations, and shipment durations) and, consequently, may estimate $\mathbf{P}(s)_i$ more accurately. To make this more concrete, consider the Markov blanket of the

MSN and BSN in Figs. 3 and 4 respectively. An important difference between these networks is that the MSN does not contain the origin (ORG) and port of loading (POL), while in contrast, the BSN does contain several binary variables representing specific locations crossed at these positions in the itinerary. This example demonstrates that, although a good may be independent of a position in the itinerary, there might still exist dependencies between the good and specific locations at the position. By binarizing all locations, we can learn these dependencies and estimate the presence of goods more accurately.

We should note that binarizing all variables in a BSN may cause some of its conditional probability tables to be structurally incomplete. The reason for this problem is that the locations of an itinerary position are mutually exclusive. Only one location can be crossed at an itinerary position. Moreover, a shipment can have only one shipment duration. When, for example, a good is conditioned on two locations that are crossed at the same itinerary position, then this probability is undefined according to Maximum Likelihood. This problem does, however, not affect the estimates of **P** that we deduce from the network because configurations involving conflicting mutual exclusive variables will not occur in the data. We may specify a prior for each probability of the BSN to avoid undefined entries in the conditional probability tables.

3.6. Derivation of a discriminative model from a Bayesian network

A BN is a particular type of generative model. It estimates the joint probability distribution $P(X)$ of X, and in turn, can be applied to infer $P(x_i | x_{j \neq i})$ indirectly by Bayes' theorem. $P(x_i | x_{j \neq i})$ can also be estimated by a probabilistic discriminative model. A probabilistic discriminative model estimates $P(x_i | x_{j \neq i})$ directly. It has been shown that this approach tends to give more accurate estimates in practice (Ng & Jordan, 2002; Roos, Wettig, Grünwald, Myllymäki, & Tirri, 2005).

We construct a set of discriminative sub-models, one for each good, that each estimate $\mathbf{P}(s)_i$ using Eq. (6). These sub-models are derived from the topology of the MSN or BSN. We do this in two steps. First, we determine the Markov blanket of each good in the shipment network. Then, we construct a discriminative sub-model for each good that predicts the presence of the good based on the variables in its Markov blanket.

The sub-models that we derive in this way from an MSN can be unnecessarily complex. Because many discriminative models, like logistic regression or neural networks, require numerical inputs, we have to binarize all features. This requirement yields sub-models that are constructed on many features that are irrelevant to predict the good under consideration. Consider again the Markov blanket of the MSN in Fig. 3. Here, $l_{\mathrm{DES}} \in \mathrm{MB}(g_{39})$ would result in 145 binary location features of which only 14 are relevant to predict g_{39}.

In contrast, a BSN models individual (binary) locations and can filter out locations that are irrelevant to predict a good. This feature yields sub-models that are much less complex and faster to train. For this reason, we only consider discriminative models that are derived from the topology of a BSN. We discuss two variations, based on logistic regression and multi-layer perceptron networks,[9] which we will refer to as BSN-LR and BSN-NN respectively.

3.6.1. BSN-LR

BSN-LR models the presence of goods on the cargo list of a shipment by a set of logistic regression models. The regression

---

[6] For an extensive treatment of Bayesian networks, see Koller and Friedman (2009).

[7] When modeling itineraries of variable length, each variable l_j includes an additional state to denote that no location is crossed at the jth position in the itinerary.

[8] Similarly, the network may contain an additional variable l_j^0 for each jth position of the itinerary to denote that no location is crossed at this position.

[9] For an extensive treatment of logistic regression and neural networks, see Bishop (1995).
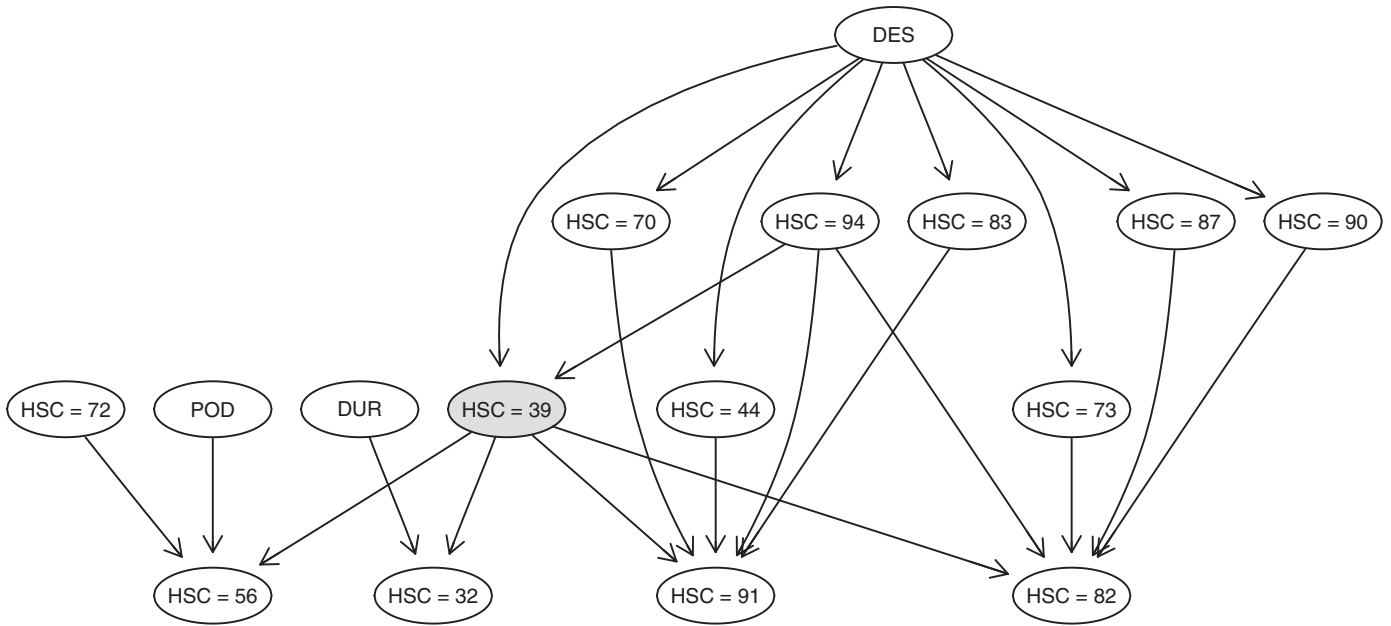
Fig. 3. The Markov blanket corresponding to a good with HS-code 39 (plastics and articles thereof) in a MSN estimated from real-world shipment data. The network consists of a set of binary nodes (HSC) representing goods, four multinomial itinerary positions (origin (ORG), port of loading (POL), port of discharge (POD) and destination (DES)), and a multinomial node (DUR) representing the shipment duration. Notice that not all of these nodes are shown in the Markov blanket depicted.
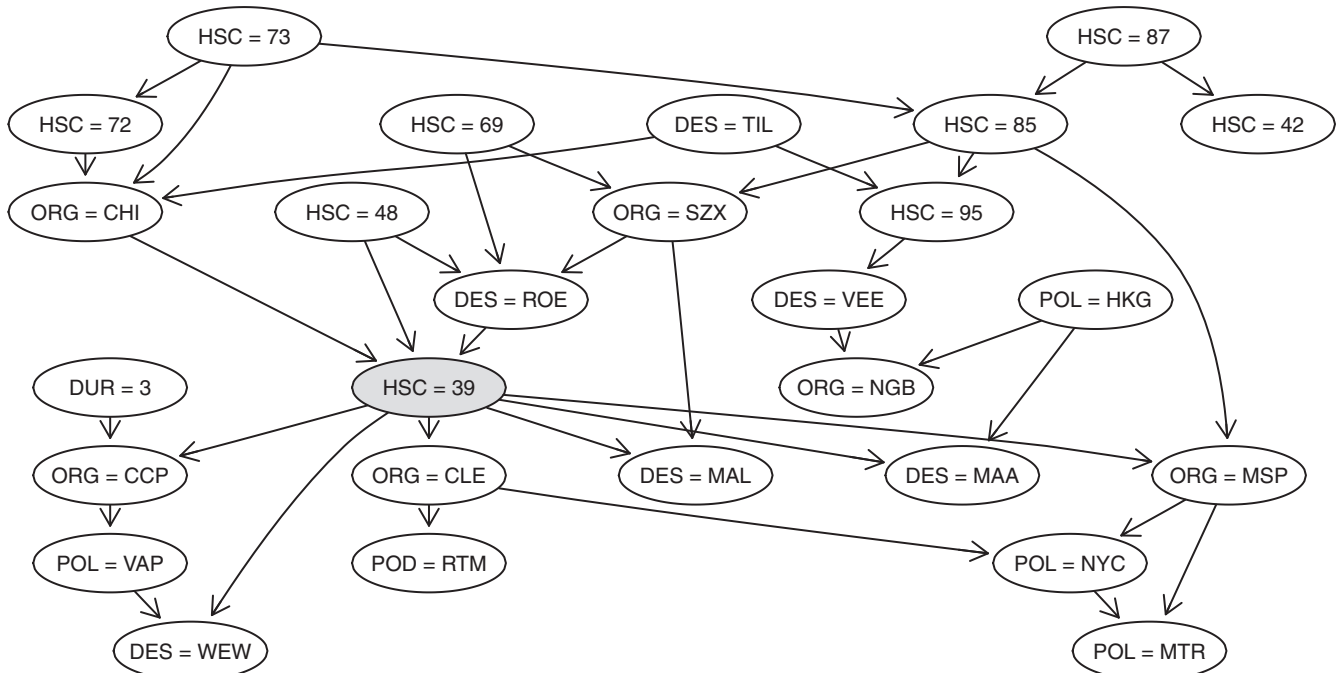

Fig. 4. The Markov blanket corresponding to a good with HS-code 39 (plastics and articles thereof) in a BSN estimated from real-world shipment data. The network is estimated from the same data as the MSN in Fig. 3 except that all nodes are binarized. Some nodes have been deliberately removed to make the graph easier to read.

model of a single good g_i can be defined as:

$$\mathbf{P}(s)_i = \sigma\left(\sum_{j=1}^{m_i} w_j \cdot MB(g_i)_j + b\right) \qquad (7)$$

where, m_i is the number of features in $MB(g_i)$, $MB(g_i)_j$ is the jth element of the Markov blanket of g_i, $w_j \in \mathbb{R}$ is a weight, $b \in \mathbb{R}$ is a bias term, and $f(x) = 1/(1 + e^{-x})$ is the sigmoid function. The sigmoid function rescales the linear combination between zero and one. The output of the model can be interpreted as $P(g_i|MB(g_i))$.

3.6.2. BSN-NN

BSN-NN models the presence of goods on the cargo list of a shipment by a set of Multi-Layer Perceptron (MLP) networks. These MLP networks operate similarly as the logistic regression models of a BSN-LR, except they process the Markov blank of each good through multiple layers of hidden neurons.

Suppose the MLP network of good g_i consists of a single hidden layer. The activation of the kth neuron, h_k, of this layer can be

defined as:

$$h_k = f\left(\sum_{j=1}^{m_i} w_{jk}^{(1)} \cdot MB(g_i)_j + b_k^{(1)}\right) \tag{8}$$

where, $w_{jk}^{(1)}$ is the weight associated with the connection between $MB(g_i)_j$ and the kth hidden neuron, $b_k^{(1)}$ is the bias of the neuron, and $f(x)$ is an activation function, e.g. the sigmoid function or hyperbolic tangent function. Accordingly, the output of the network can be defined as:

$$\mathbf{P}(s)_i = \sigma\left(\sum_{j=1}^{l} w_j^{(2)} \cdot h_j + b^{(2)}\right) \tag{9}$$

where, l is the number of neurons in the hidden layer, $w_j^{(2)}$ is the weight associated with the connection between the jth hidden neuron and the output g_i, and $b^{(2)}$ is the bias of the output.

4. Experimental setup

In this section, we discuss a series of experiments in which the performance of the BSN and two probabilistic discriminative models were evaluated on real-world shipment data. We elaborate on the characteristics of the shipment data (Section 4.1), the model implementations (Section 4.2), and the methodology by which the performance of the models was measured (Section 4.3).

4.1. Shipment data

We extracted a sample of shipments from the supply chain repository of an international freight forwarder. The sample contains details of shipments that were transported overseas to the Netherlands between April 2012 and June 2013. It includes details of the goods that were conveyed by the shipments as specified on the import declaration, together with itinerary details that were specified on the bill of lading corresponding the cross-border transportation.

Some pre-processing was applied to prepare the data for analysis. Because of the relatively small sample size, most six-digit HS-codes were shipped only a few times. To prevent over-fitting, we extracted the first two digits (chapter codes) of the HS-codes. Furthermore, the sample included three locations that were crossed in the itinerary of the shipments: the origin (ORG), port of loading (POL), and port of discharge (POD). Data of the exact destination (DES) was not available. Therefore, we used the location of the customer who imported the goods as an approximation to where the goods were most likely shipped[10] Finally, we calculated the shipment duration from the port of loading to the port of discharge based on the ATD (Actual Time of Departure) and ATA (Actual Time of Arrival). Equal width binning (Dougherty, Kohavi, & Sahami, 1995) was applied to transform the shipment duration to a discrete feature consisting of ten approximately equally spaced time intervals.

Not all shipments in the sample could be used for model evaluation. Because of the high dimensionality of the data, some combinations of goods and trajectories are very rare. We removed these rare combinations by applying the following filter rules:

1. Goods that have been shipped less than 15 times are removed.
2. Shipments with trajectories that have been taken less than 3 times are removed.

The remaining sample included 10,149 shipments, 50 different types of goods, and 625 unique itineraries. Table 2 shows a small subset of the sample.

---

[10] We retrieved this location by querying the Google Maps API by the company name of each customer.

Table 2
Four shipments of the sample. The itinerary of each shipment consists of four locations. These include the origin (ORG), port of loading (POL), port of discharge (POD), and destination (DES). The duration represents the time elapsed between the departure at the port of loading and the arrival at the port of discharge.

Cargo list (HS-codes)	Itinerary < ORG,POL,POD,DES >	Duration (Hours)
{73, 84, 85}	< MEM,CHS,RTM,GEI >	$10.60 < x \le 19.00$
{33, 34}	< TOR,MTR,RTM,SAS >	$1.89 < x \le 10.60$
{35, 39, 87}	< MSP,MTR,ANR,BES >	$1.89 < x \le 10.60$
{39, 40, 73, 84, 85}	< CLE,NYC,OMD,TIE >	$10.60 < x \le 19.00$

4.2. Model implementation

The sample was partitioned into two separate sets for training and evaluation purposes. Approximately 75% of the shipments were sampled by stratified sampling with the itinerary as strata and put in a training set. The remaining 25% of the shipments were put in a test set.

We applied R package bnlearn (Scutari, 2010) to construct a BSN on the shipments in the training set. The structure of the network was estimated by a hill-climbing search. The search algorithm started with the mutual independence model (empty graph), then tried to find a better network by iteratively adding, removing or reversing edges, and finally stopped when no further improvements to the current network could be made. To avoid over-fitting, we scored candidate networks by:

$$\phi(\mathcal{BN}, D) = \log \mathcal{L}(\mathcal{BN}, D) - cp \tag{10}$$

where, $\mathcal{L}(\mathcal{BN}, D)$ is the likelihood function of the candidate network, p is the number of parameters of the network, and c is a penalty coefficient that controls how strongly the complexity of the network is penalized. We experimented with different penalties $c \in \{0.01, 0.009, \dots, 0.001\}$. In our case, $c = 0.002$ gave the best results. The probabilities of the network were estimated by Bayesian parameter estimation. We performed this estimation procedure with a Beta(5, 5) distribution as prior for each probability of the network. Inference in the network was performed by likelihood-weighting (Fung & Chang, 1989).

Moreover, we derived a BSN-LR and BSN-NN from the BSN according to the procedure described in Section 3.6. The regression models of BSN-LR were constructed by the glm function in R package stats (R Core Team, 2013). It estimated the weights and bias terms of the models by the iterative re-weighted least square algorithm (Nelder & Wedderburn, 1972). The neural networks of BSN-NN were constructed by R package nnet (Venables & Ripley, 2002) and contained one hidden layer with sigmoid activations. The weights and bias terms of the networks were estimated by minimizing the cross-entropy using the BFGS algorithm in conjunction with back-propagation (Werbos, 1982).

The neural networks of BSN-NN have some hyper-parameters that need to be tuned. The most important ones are the number of neurons l in the hidden layer and the amount of weight decay λ. We tuned these parameters by performing holdout cross-validation (Kohavi, 1995) with R package caret (Kuhn, 2008). During the cross-validation procedure, approximately 10% of the shipment in the training set were randomly removed and put in a separate holdout set. Accordingly, a set of neural networks were constructed on the remaining training set having a different number of hidden neurons $l \in \{10, 20, 30, 40\}$ and weight decay $\lambda \in \{10^{-2}, 10^{-3}\}$. The classification accuracy of these networks was evaluated on the holdout set. The configuration that achieved the highest accuracy for each good on the holdout set was selected.

Table 3
The elements of a confusion matrix for a binary classification problem.

		Actual	
		True (1)	False (0)
Model / Audit	True (1)	*True Positive Rate (TPR)*	*False Positive Rate (FPR)*
	False (0)	*False Negative Rate (FNR)*	*True Negative Rate (TNR)*

4.3. Evaluation metrics

The sample was unlabeled and did not contain information about which shipments involved document fraud. Therefore, we evaluated the models by generating artificial fraud incidents and determining how many incidents the models were able to detect. We generated these fraud incidents in a separate miscoding set and smuggling set. The miscoding set and smuggling set are identical to the test set except that they contain a small random subset of approximately 10% of the shipments whose cargo list is artificially manipulated. Miscoding in the miscoding set was generated by randomly adding a good to the cargo list of a shipment. Similarly, smuggling in the smuggling set was generated by randomly removing a good from the cargo list of a shipment.

We evaluated the ability of the models to distinguish between shipments with an original cargo list and those with a manipulated cargo list by constructing a confusion matrix. Table 3 shows a confusion matrix for this binary classification problem. It consists of two rows and two columns. The rows of the matrix denote the cases were the model produced an alarm or not, whereas the columns denote the cases were shipments involved fraud or not. Given these cases, the matrix defines four prediction rates.

The prediction rates of the confusion matrix are computed as follows. Suppose $A(s)$ and $F(s)$ are defined as:

$$A(s) = \begin{cases} 1 & \text{if an alarm is produced for } s \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

$$F(s) = \begin{cases} 1 & \text{if } s \text{ is fraudulent} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

Then, for model m we have:

$$TPR_m = \frac{1}{|D|} \sum_{s \in D} A(s)F(s) \tag{13}$$

$$FPR_m = \frac{1}{|D|} \sum_{s \in D} A(s)(1 - F(s)) \tag{14}$$

$$FNR_m = \frac{1}{|D|} \sum_{s \in D} (1 - A(s))F(s) \tag{15}$$

$$TNR_m = \frac{1}{|D|} \sum_{s \in D} (1 - A(s))(1 - F(s)) \tag{16}$$

We estimated these prediction rates ten times while each iteration had a different partition of shipments into training and testing examples, and different incidents in the miscoding set and smuggling set. Accordingly, we computed the average prediction rates.

From these average prediction rates, we derived the average precision, recall and F_1. Precision and recall are defined as (Olson & Delen, 2008):

$$\text{Precision} = \frac{TPR}{TPR + FPR} \tag{17}$$

$$\text{Recall} = \frac{TPR}{TPR + FNR} \tag{18}$$

Precision is the fraction of correctly identified shipments containing miscoding or smuggling compared to all shipment with miscoding or smuggling. Recall is the fraction of shipments for which a correct alarm was produced. Combining both measures yields the F_1 score (Olson & Delen, 2008):

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{19}$$

The F_1 score is the harmonic mean of precision and recall. It constitutes a single measure to evaluate the performance of a set of competing models.

Besides the data-driven models, we also measured the performance of a set of random audits which generated the same number of alarms as the models but produced alarms randomly. This procedure allows us to compare the models with the case of applying random audits to detect document fraud. The expected values for the prediction rates of these random audits can be easily derived. For example, the expected true positive rate of random audit r is:

$$\mathbb{E}(TPR_r) = \mathbb{E}\left(\frac{1}{|D|} \sum_{s \in D} A(s)F(s) \right) \tag{20}$$

$$= \frac{1}{|D|} \sum_{s \in D} P_A P_F \tag{21}$$

$$= P_A P_F \tag{22}$$

Here, $A(s)$ and $F(s)$ are replaced by respectively the alarm rate P_A and fraud rate P_F of the random audit. Similarly, we have:

$$\mathbb{E}(TPR_r) = P_A P_F \qquad \mathbb{E}(FPR_r) = P_A(1 - P_F) \tag{23}$$

$$\mathbb{E}(FNR_r) = (1 - P_A)P_F \qquad \mathbb{E}(TNR_r) = (1 - P_A)(1 - P_F) \tag{24}$$

Furthermore, from Eqs. (17), (18), and (19) it follows that:

$$\mathbb{E}(\text{Precision}) = P_F \qquad \mathbb{E}(\text{Recall}) = P_A \qquad \mathbb{E}(F_1) = 2 \cdot \frac{P_F \cdot P_A}{P_F + F_A} \tag{25}$$

We compared the average precision, recall, and F_1 of the models with a random audit that generates the same number of alarms. We did this by setting the alarm rate of the random audit equal to the average alarm rate of the corresponding model, i.e. $P_A = \bar{TPR}_m + \bar{FPR}_m$. Fraud rate $P_F = \bar{TPR}_m + \bar{FNR}_m \approx 0.1$ is constant in all our experiments.

5. Results

Tables 4 and 5 summarize the results of the experiments. All experiments were performed with a risk threshold of $\alpha = 0.1$ and $\beta = 0.9$. These thresholds imply that the models produced miscoding alerts for goods which were listed on the cargo list but had an probability of being present 10% or less. Moreover, they produced smuggling alerts for goods which were not listed on the cargo list but had an probability of being present 90% or higher.

Overall, the results show that the data-driven models provide fraud alarms of much better quality than the corresponding random audits. The models achieved consistently higher F_1 scores.

Table 4
The results of the miscoding experiments averaged over ten evaluation iterations.

	Miscoding					
	BSN	Random audit	BSN-LR	Random audit	BSN-NN	Random audit
Precision	0.3469	0.0999	0.3803	0.0999	0.4383	0.0999
Recall	0.9920	0.2857	0.9858	0.2589	0.9923	0.2262
F1	0.5140	0.1480	0.5489	0.1442	0.6080	0.1386

Table 5
The results of the smuggling experiments averaged over ten evaluation iterations.

	Smuggling					
	BSN	Random audit	BSN-LR	Random audit	BSN-NN	Random audit
Precision	0.5094	0.0999	0.5463	0.0999	0.6073	0.0999
Recall	0.6847	0.1343	0.5234	0.0957	0.8854	0.1456
F1	0.5842	0.1146	0.5346	0.0977	0.7204	0.1185

Moreover, the precision and recall reveal that the models are on average far more likely to generate a correct alarm and detect a considerably larger portion of the fraud incidents in the miscoding set and smuggling set. Consider the results of the BSN. The model has an average precision and recall of respectively 35% and 99% for miscoding, and 51% and 69% for smuggling. In contrast, the random audit that generates the same number of alarms as the BSN achieved only an average precision and recall of respectively 10% and 29% for miscoding, and 10% and 13% for smuggling.

Closer examination of the results also reveals that BSN-NN performed even better than the BSN. BSN-NN achieved an average precision and recall of respectively 44% and 99% for miscoding. In contrast, the BSN achieved the same recall but with lower precision. The alarms for smuggling are of slightly better quality. BSN-NN achieved an average precision and recall of respectively 61% and 89% for smuggling. Again, this is better than the precision and recall of the BSN. These results seem to be in line with earlier work of Ng and Jordan (2002) and Roos et al. (2005) who showed that discriminative models typically give better performance than generative models.

6. Conclusions

We conclude from our experiments that intelligent fraud detection systems can considerably improve the detection of miscoding and smuggling in the international shipping industry. By leveraging the shipment data generated by supply chains, these systems can make a better selection of shipments that require further fraud analysis than random audits. Our results suggest they are on average far more likely to select a fraudulent shipment and overall detect a much more significant portion of the fraud cases. This observation indicates that intelligent fraud detection systems are an important addition to the risk management practices of shipping companies and customs authorities.

Regarding the design of the fraud detection system, there is a trade-off between model performance and flexibility. Discriminative models tend to perform better than generative models. In our experiments, BSN-NN outperformed the BSN on both miscoding and smuggling. However, a drawback of discriminative models is that they usually cannot handle incomplete data, while this is an important requirement in this particular application. Shipment documentation consists of a set of documents that are collected when a shipment moves through the supply chain. Therefore, the documentation is in most cases only complete when a shipment has already passed the customs borders of the destination country, and physically inspecting the cargo is no longer possible. In contrast, generative models, such as Bayesian networks, can deal with missing data very well and perform fraud detection at any stage of the supply chain.

We recognize that our method of generating artificial fraud incidents is somewhat oversimplified. In practice, miscoding and smuggling are typically committed in a more sophisticated matter than merely adding or removing a random good from the cargo list of a shipment. Unfortunately, it is difficult to measure the extent to which fraud detection systems can detect real fraud cases. Customs authorities are reluctant to share any data about which shipments turned out to be fraudulent because of privacy and security reasons. Even if such data would be available, then it would probably not include a correct label for each shipment because customs authorities cannot physically inspect each shipment that crosses the borders of a country. Still, it would be interesting to investigate how intelligent fraud detection systems would perform in this case. We leave this open for future research.

Moreover, in future research, our work can be further improved in several aspects. First, we predict miscoding and smuggling based on the goods conveyed by a shipment and basic information of its itinerary. Better predictions may be obtained when modeling more details about goods, like their prices and weights, and the itinerary. Second, we evaluated our models on data of shipments that were shipped to the Netherlands and for which only four itinerary locations were available. Future research should evaluate if the same results are obtained on global shipment data containing more detailed itineraries.

References

Aronson, J. E., Liang, T.-P., & Turban, E. (2005). *Decision support systems and intelligent systems*. Pearson Prentice-Hall.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford university press.

Camossi, E., Dimitrova, T., & Tsois, A. (2012). Detecting anomalous maritime container itineraries for anti-fraud and supply chain security. In *European intelligence and security informatics conference (eisic)* (pp. 76–83). IEEE.

Chahuara, P., Mazzola, L., Makridis, M., Schifanella, C., Tsois, A., & Pedone, M. (2014). Inferring itineraries of containerized cargo through the application of condi-

tional random fields. In *Ieee joint intelligence and security informatics conference (jisic)* (pp. 137–144). IEEE.

Cowell, R. G., Dawid, P., Lauritzen, S. L., & Spiegelhalter, D. J. (2006). *Probabilistic networks and expert systems: exact computational methods for bayesian networks*. Springer Science & Business Media.

Digiampietri, L. A., Roman, N. T., Meira, L. A., Ferreira, C. D., Kondo, A. A., Constantino, E. R., et al. (2008). Uses of artificial intelligence in the brazilian customs fraud detection system. In *Proceedings of the 2008 international conference on digital government research* (pp. 181–187). Digital Government Society of North America.

Dimitrova, T., Tsois, A., & Camossi, E. (2014). Development of a web-based geographical information system for interactive visualization and analysis of container itineraries. *International Journal of Computer and Information Technology, 3*(01), 1–8.

Dougherty, J., Kohavi, R., & Sahami, M. (1995). Supervised and unsupervised discretization of continuous features. In *Machine learning: Proceedings of the twelfth international conference: 12* (pp. 194–202).

Ezawa, K. J., & Schuermann, T. (1995). Fraud/uncollectible debt detection using a Bayesian network based learning system: A rare binary outcome with mixed data structures. In *Proceedings of the eleventh conference on uncertainty in artificial intelligence*. In *UAI'95* (pp. 157–166). San Francisco, CA, USA

Feelders, A. (2012). A new parameter learning method for Bayesian networks with qualitative influences. arXiv:1206.5245.

Filho, J. J., & Wainer, J. (2007). Using a hierarchical Bayesian model to handle high cardinality attributes with relevant interactions in a classification problem. In *International joint conference on artificial intelligence IJCAI* (pp. 2504–2509).

Fung, R., & Chang, K.-C. (1989). Weighing and integrating evidence for stochastic simulation in Bayesian networks. In *Proceedings of the fifth conference on uncertainty in artifical intelligence (uai-89)* (pp. 475–482).

Gordhan, P. (2007). Customs in the 21st century. *World Customs Journal, 1*(1), 49–54.

Hesketh, D. (2010). Weaknesses in the supply chain: Who packed the box. *World Customs Journal, 4*(2), 3–20.

Hill, G. N., & Hill, K. (2009). *Nolo's plain-english law dictionary*. Nolo.

Hua, Z., Li, S., & Tao, Z. (2006). A rule-based risk decision-making approach and its application in China's customs inspection decision. *Journal of the Operational Research Society, 57*(11), 1313–1322.

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications, 32*(4), 995–1003.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International joint conference on artificial intelligence IJCAI: 14* (pp. 1137–1145).

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. MIT press.

Kuhn, M. (2008). Building predictive models in r using the caret package. *Journal of Statistical Software, 28*(1), 1–26.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General), 135*(3), 370–384.

Ng, A., & Jordan, M. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In *Advances in neural information processing systems 14* (pp. 841–848).

Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques*. Springer Science & Business Media.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

R Core Team (2013). R: *A language and environment for statistical computing*. R Foundation for Statistical ComputingVienna, Austria.

Roos, T., Wettig, H., Grünwald, P., Myllymäki, P., & Tirri, H. (2005). On discriminative Bayesian network classifiers and logistic regression. *Machine Learning, 59*(3), 267–296.

Scutari, M. (2010). Learning Bayesian networks with the bnlearn r package. *Journal of Statistical Software, 35*(3), 1–22.

Taniguchi, M., Haft, M., Hollmén, J., & Tresp, V. (1998). Fraud detection in communication networks using neural and probabilistic methods. In *Proceedings of the 1998 ieee international conference on acoustics, speech and signal processing: 2* (pp. 1241–1244). IEEE.

Triepels, R., Feelders, A., & Daniels, H. (2015). Uncovering document fraud in maritime freight transport based on probabilistic classification. In *Computer information systems and industrial management*. In *Lecture Notes in Computer Science: 9339* (pp. 282–293). Springer.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (4th ed.). New York: Springer.

Villa, P., & Camossi, E. (2011). A description logic approach to discover suspicious itineraries from maritime container trajectories. In *Geospatial semantics*. In *Lecture Notes in Computer Science: 6631* (pp. 182–199). Springer.

Werbos, P. J. (1982). Applications of advances in nonlinear sensitivity analysis. In *System modeling and optimization* (pp. 762–770). Springer.

World Customs Organization (2012). Illicit trade report.

World Shipping Council (2004). Trade statistics. http://www.worldshipping.org/about-the-industry/global-trade/trade-statistics.

Yaqin, W., & Yuming, S. (2010). Classification model based on association rules in customs risk management application. In *International conference on intelligent system design and engineering application (isdea): 1* (pp. 436–439). IEEE.