Carancho - A Decision Support System for Customs

Norton Trevisan Roman[1], Everton Rufino Constantino[1], Helder Ribeiro[1], Jorge Jambeiro Filho[1,2], Antonella Lanna[2], Siome Klein Goldenstein[1], and Jacques Wainer[1]

[1] Institute of Computing – UNICAMP – Brazil
[2] Brazil's Federal Revenue

Abstract. This paper discusses Carancho, a decision support system for fraud detection in customs transactions. The system is based on the outlier detection paradigm as a form of detecting frauds. It has both an interactive and a batch mode. In the interactive mode, a customs expert simultaneously decides on whether a particular transaction is an outlier, and what is a "median" transaction against which the outlier is contrasted. In the batch mode, the system only selects potential outliers for further inspection. The novelty of the system is that instead of attacking the outlier problem in the many dimensions, we project the data into a set of specially formulated relevant dimensions, and detect the outliers in those dimensions.

1 Introduction

Foreign commerce has historically been of great importance as an economical and political instrument worldwide. Tax policy, among other instruments, are forms by which a government controls the trade of goods and services. The problem is that, as one might expect, whenever there is someone charging taxes, there is also someone else trying to avoid paying them.

Unfortunately, this is not only a simple matter of tax evasion. According to the U.S. Department of State [1], practises such as commodities overvaluation allow corporations to perform fund transfers transparently, possibly concealing money laundering schemes. These activities can be directly connected to drug traffic and smuggling. Under this light, an apparently minor tax-evasion offence could be just the tip of the iceberg of more severe crimes.

Some of the common problems found on customs transactions are:

- *Over/Undervaluation*: As mentioned before, incorrect price estimation can conceal illicit transfers of funds. Money laundering schemes in the case of overvaluation, and tax evasion offences in the case of undervaluation.
- *Classification errors*: When assigning a product to one of the predefined categories, the importer can make a honest mistake, misclassifying it. Nevertheless, such a mistake could also conceal an instance of tax evasion, should the misclassification lead to lower tax charges.

– *Origin errors*: Such errors happen when an importer incorrectly declares the goods' country of origin. This is a direct effect of special customs restrictions regarding particular combinations of goods and their origin.
– *Smuggling*: Sometimes importers smuggle different materials into the country amongst the goods they are actually importing. This problem, as reported in [1], could also conceal money laundering schemes, specially when the smuggled material is a high-valued commodity, such as gold.

In this paper, we describe *Carancho*, a graphical decision support system designed to help customs officers decide what should be inspected taking into account all the past international operations.

2 Approaching the Problem

To address the four basic issues stated in the previous section, we follow some of the general ideas regarding the use of unsupervised learning for fraud detection presented in [2]. More specifically, we address such questions as outlier detection problems.

Our choice for the outlier detection framework is based on the assumption that the majority of international trading operations are in order, *i.e.*, that most importers have not only followed the laws properly, but also made no mistakes in the process of importing goods and/or services.

The central problem in applying standard outlier detection techniques to this problem is that the number of relevant attributes for an import declaration is extremely large. Although some of these attributes be continuous, such as weight and price, others are categorical with sometimes a rather large number of classes. Take as an example the merchandise code, that can assume one of approximately ten thousand different types, or the country of origin, which can be one of a set of over a hundred values.

Instead of tackling the outlier detection problem in a large dimensional space, we followed a different approach. Customs experts we interviewed declared that they felt that for each different type of fraud, a small set of single dimensions were enough to detect the outlier with some confidence. Regarding over and under evaluation, the experts defined a set of single dimensions[3] which they felt were the most significant. Their exact definitions are considered confidential, thus this strategic information cannot be made public in this paper.

It is important to notice that the officers' experience and expertise are codified into the system through these dimensions. We think that this approach, besides allowing outlier detection algorithms to be used, gives the system a higher adaptability to new situations than we could have achieve by hard coding a set of rules, as suggested in [3].

We followed a double approach in this research. The Carancho system can work as both a user directed decision support system, which helps the expert to select both outliers and medians, or as a more autonomous batch search for the outliers. Both approaches are described in more detail below.

[3] Currently we are dealing only with three of them.

2.1 Graphical-Assisted Manual Selection

The manual selection approach relies on the expertise and experience of the customs officer to determine what could be considered a suspicious import. Within this framework, the officer is responsible for deciding about the inspection of some goods based on a graphical representation of all operations which are similar to the one under evaluation.

Inspectors can define the level of similarity on the fly, applying filters to the database of transactions. The program retrieves the filtered relevant information, groups the data according to a set of predefined dimensions, and plots their histograms. This is a poor man's attempt to display the transactions' distribution according to these dimensions. The inspector can then, based on this overview of all the similar transactions, decide whether to inspect the cargo more carefully or to clear it right away.

2.2 Batch Search

The batch search is appropriate for inspection of all operations that have already been cleared, but that may deserve later investigation on documentation and financial statements. Basically, the system processes all the database, looking for outliers. This search is an attempt to automate the manual approach, using standard statistical estimation techniques.

Outliers detection takes place as follows. The system searches the whole database, grouping import documents according to their product code[4]. It then analyses each group according to the aforementioned dimensions.

For each of the strategic dimensions, the system finds its robust mean and standard deviations. Then, it looks for transactions that are more than three standard deviations away from the mean in any of these dimensions. That is, it defines a dimension outlier as any import operation whose declared value, for a specific dimension d, is

$$value(d) \geq (mean(d) + 3 \times \sigma(d))$$

An outlier is defined as any import that is an outlier in any of the dimensions. Or, in other words

$$outlier = outlier(d1) \vee outlier(d2) \vee outlier(d3)$$

The results of outlier detection are sent to a file, for later analysis.

3 Results

We ran the batch mode of the system on a 1.5 million transactions database. The outlier detection rate was around 0.9%. That is, from the total amount of import operations, around 14000 were considered outliers.

[4] Similar to the Harmonised System, developed by the World Customs Organisation.

This is a very interesting result, for it allows the customs department to pay special attention to these "strange" operations, as opposed to deciding which ones should be cleared and which should be inspected on a more subjective basis. This saves time and effort, not to mention financial resources.

Naturally, we have to bear in mind that being an outlier does not imply being an outlaw. Some outlier operations are in fact proper ones, they simply do not fit the pattern followed by the majority of similar operations. Nevertheless, this system will increase the customs officers' efficiency, making the decision process swifter and speeding the customs process.

We are still working on assessing the precision of our system. We choose not to use recall as an evaluation measure because there is no way of telling which irregular operations escaped detection.

4 Conclusion

In this paper we presented Carancho – a decision support system for the customs department, aimed at detecting some types of frauds.

We proposed a new approach to the problem, set in two fronts. The first front is intended to help customs officers to make their decisions more quickly, by showing them a graphical representation of the distribution of similar import operations in the past.

The second front is intended to detect those operations that are very unusual but, for some reason, were cleared from physical inspection. In sum, the first approach is meant to detect outliers as they happen, on the fly, whereas the second one is meant to detect past outliers.

Another novelty of our model is the fact that we broke up a multi-dimensional vector space (the dimensions analysed by the system) into a set of one-dimensional spaces, so that we could look for outliers in each of them separately, and then combine the results. Despite the fact that this measure simplified our task, we still need to determine how good a decision it was.

References

1. U.S. Department of State: International Narcotics Control Strategy Report. Volume II: Money Laundering and Financial Crimes. (2005)
2. Bolton, R.J., Hand, D.J.: Statistical fraud detection: A review. Statistical Science **17**(3) (2002) 235–255
3. Singh, A.K., Sahu, R.: Decision support system for hs classification of commodities. In: Proceedings of the 2004 IFIP International Conference on Decision Support Systems (DSS 2004). (2004)