

## CONSTRUCTING DECISION TREE BASED ON QUESTIONNAIRES TO DETECT A POSSIBLE CORRUPTION IN LOGISTICS

Dmitry Stefanovskiy, Sergey Maruev, Xavier Tejada

**Abstract:** *Decision trees are the well-known a popular tool in the areas where non-numerical data are used. In the paper we demonstrate the application of decision trees for analysis of International logistics. First of all we cluster data to reduce the number of variants for comparison. Then the regression analysis is completed and the significant parameters combinations are determined. The logical function and decision tree could be useful for revealing possible corruption schemes in logistics. Finally we complete experiments with this decision tree. The suggested method can be used for constructing a decision tree based on a large amount of similar data.*

**ACM Classification Keywords:** 1.2 Artificial Intelligence

**Keywords:** *decision trees, logistics, corruption*

---

### Introduction

---

The activity of one of the authors is related with a product transportation from West Europe to East Europe. It was detected the essential changes in transportation flows during recent years and this fact caused our interest to the topic. Our goal is to construct a classifier for testing hypothesis about possible corruption in logistics.

There are many newspaper and journal articles and books discussing various aspects of corruption. But only few of them contain its formal models. That is due to the lack of numeric data to be processed.

Typical approaches to revealing hidden corruption use analysis of numerical characteristics reflecting prices of goods, taxes, etc. [Levin, 2011; Rose-Ackerman, 1978] Same works discuss the behavior of private and state agents in different macroeconomic conditions [Shleifer, 1993; Olimpieva, 2007]. Our approach is based on qualitative information we get from the companies dealing with product transportation from West Europe to East Europe. In the paper this information is presented in the form of questionnaires. After processing the information is transformed to a decision tree. This decision tree can be a detector for revealing simple cases related to corruption.

In section 2 we describe a questionnaire. In section 3 we form and analyze a decision table. Here we construct a decision tree itself. In section 4 we discuss the experiment. Section 5 contains the conclusion

---

### Questionnaire and data preprocessing

---

A questionnaire was developed. The questions were offered to managers of transportation companies on conditions of anonymity. In 3 months about 5 thousands records in a database were received. Each record stores combinations of parameters (answers) corresponding to one transportation. The questionnaire includes the following questions:

- 1) Is the cargo intended for the Russian Federation? (1-Yes, 0-No)
- 2) Is the date of sale of the cargo the same as the date it crossed the border? (1-Yes, 0-No)
- 3) Do you face corruption at the customs on the borderline? (1-Yes, 0-No)
- 4) Have any companies registered in Byelorussia become your clients since January, 1, 2012? (1-Yes, 0-No)

- 5) Is the recipient registered in the Russian Federation? (1-Yes, 0-No)
- 6) Do you participate in the corruption at the customs of the Russian Federation? (1-Yes, 0-No)
- 7) Do you participate in the corruption in case herbal and veterinary control of the Russian Federation? (1-Yes, 0-No)
- 8) Do you participate in the corruption at the customs of the EU? (1-Yes, 0-No)
- 9) Do you participate in the corruption in case herbal and veterinary control of the EU? (1-Yes, 0-No)
- 10) Where was the border crossed? (Poland-1, Baltic States-0)
- 11) Do you think that the official salary of customs officials in the Russian Federation should be higher? (1-Yes, 0-No)
- 12) Do you think that the official salary of customs officials in the EU should be higher? (1-Yes, 0-No)
- 13) Do you think you will lose more if you participate in corruption activities at the EU customs? (1-Yes, 0-No)
- 14) Do you think you will lose more if you participate in corruption activities at the Russian customs? (1-Yes, 0-No)
- 15) Do you think you will lose more if you participate in corruption activities in case herbal and veterinary control of the Russian Federation? (1-Yes, 0-No)
- 16) Do you think you will lose more if you participate in corruption activities in case herbal and veterinary control of the EU? (1-Yes, 0-No)

The preliminary analysis showed that many records had similar sets of values of parameters. To group them we used k-means method with the selection of the best k on the basis of Dunn index

Having determined the cluster centers it proved that there were parameters with the almost identical values in all the centers. Excluding these parameters let us simplify further analysis of the data. Sets of parameters of the cluster centers are shown in Table 1. Here each parameter  $x_i$  is related with the question-i from the questionnaire.

Table 1. Sets of parameters of the cluster centers

Cluster	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$	$x_{13}$	$x_{14}$	$x_{15}$	$x_{16}$
1	1	1	1	1	0	1	0	1	0	0	1	1	0	0	0	0
2	0	1	1	1	0	1	0	1	0	0	1	1	0	0	0	0
3	0	1	1	1	0	1	1	1	0	0	1	1	0	0	1	0
4	1	0	0	1	0	1	0	0	0	1	1	1	0	0	0	0
5	1	0	0	1	0	1	0	0	0	1	1	1	0	0	0	1
6	1	1	1	1	0	1	0	1	0	0	1	1	0	0	1	0
7	1	1	1	1	0	1	0	1	1	0	1	1	0	0	0	0
8	1	1	1	0	1	1	0	1	0	0	1	1	0	0	0	0
9	1	1	1	1	1	1	0	0	0	0	1	1	1	0	0	0
10	0	0	1	1	0	1	1	0	0	1	1	1	0	0	0	0
11	1	1	1	1	0	1	0	1	1	0	1	1	0	0	0	0
12	1	0	0	1	0	1	0	1	0	0	1	1	0	0	0	0
13	1	1	1	1	1	1	0	0	0	0	1	1	0	1	0	0
14	0	1	1	1	0	1	1	0	0	1	1	1	0	0	0	0
15	1	1	1	1	0	1	1	1	0	1	1	1	0	0	0	1

For the further analysis the parameters  $x_2$ ,  $x_7$ ,  $x_8$  and  $x_{10}$  are chosen.

## Constructing decision tree

Let's take  $x_{10}$  as a dependent variable. The regression analysis including the individual variables and their pair productions allowed to construct the following formula:

$$x_{10} = a_0 + a_1x_2 + a_2x_7 + a_3x_8 + a_4x_2x_8 + a_5x_2x_7 + a_6x_8x_7$$

Coefficients and estimators are shown in table 2.

Table 2. Coefficients of the regression equation

		Estimate	Std. Error	t value
$a_0$	(Intercept)	1.00E+00	1.754E-16	5.700E+15
$a_1$	dust_data[, 2]	-1.00E+00	2.481E-16	-4.031E+15
$a_2$	dust_data[, 8]	3.694E-16	3.038E-16	-3.291E+15
$a_3$	dust_data[, 7]	1.00E+00	3.038E-16	1.216E+00
$a_4$	dust_data[, 2]:dust_data[, 8]	1.00E+00	3.652E-16	2.738E+15
$a_5$	dust_data[, 2]:dust_data[, 7]	1.00E+00	4.297E-16	2.327E+15
$a_6$	dust_data[, 2]:dust_data[, 7]	-1.00E+00	4.051E-16	-2.468E+15

According the regression equation a logical function is constructing in the form:

$$x_{10} = \text{xor}(\text{xor}(\text{xor}(\text{xor}(x_2, !x_8), (x_2 \& x_8)), (x_2 \& x_7)), !(x_7 \& x_8)),$$

xor, &, ! - the logical operations: exclusive disjunction, conjunction and negation.

Fig.1 illustrates the decision tree related with the logical function

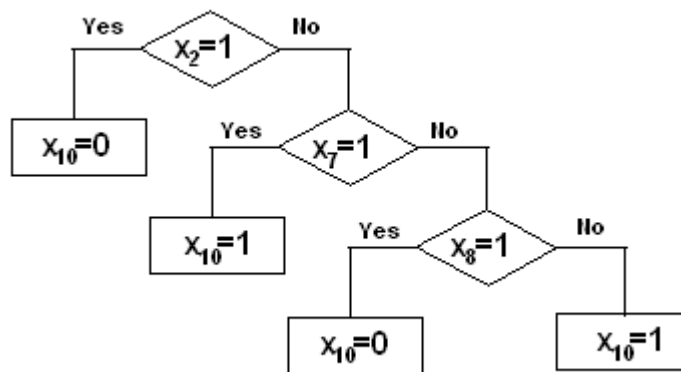


Fig.1. Logical tree

This study argues that the choice of a route can indicate a possibility to participate in a corruption scheme. Statistical verification confirms this statement.

## Conclusion

In the paper the decision tree based on questionnaire was constructed. The clustering allowed to reveal the significant combinations of parameters and to reduce the number of variants for comparison. The regression analysis determined the significant combinations of parameters for construction of logic function and the decision tree.

This decision tree was used for analysis of transportation flows of goods from West Europe to East Europe. The experiment with the decision tree was completed and the results showed good coincidence with statistical data. The proposed method can be used for constructing a decision tree based on a large amount of similar data.

---

## Bibliography

---

- [Levin, 2011] Levin M.I., Levina E.A., Pokatovich E.V. Lectures on the Economics of Corruption. M.: HSE, 2011. (rus)
- [Olimpieva, 2007] I. Olimpieva. Fighting windmills? Socio-anthropological approach to the investigation of corruption / Comp. and holes. Ed. IB Olimpieva, O. Pachenkov. Aletheia, St. Petersburg., 2007. (rus)
- [Rose-Ackerman, 1978] Rose-Ackerman S. Corruption: a Study in Political Economy. New York; San Francisco; London: Academic Press, 1978.
- [Shleifer, 1993] Shleifer A. and Vishny R.W. Corruption. The Quarterly Journal of Economics., 107 N 33, 1993.

---

## Authors' Information

---



**Dmitry Stefanovskiy** - Assoc. Prof., Ph.D, the Russian Presidential Academy of national economy and public administration; Prosp. Vernadskogo 82, bld. 1, Moscow, 119571, e-mail: [dstefanovskiy@gmail.com](mailto:dstefanovskiy@gmail.com)

*Major Fields of Scientific Research: mathematical modelling, world economy*



**Sergey Maruev** – Professor, The Russian Presidential Academy of National Economy and Public Administration, Prosp. Vernadskogo 82, bld. 1, Moscow, 119571, Russia; e-mail: [maruev@gmail.com](mailto:maruev@gmail.com)

*Major Fields of Scientific Research: mathematical modeling, economics of education, business process modeling*



**Javier Tejada Cárcamo** – Professor of Computer Science Department, San Pablo Catholic University; Research and Software Development Center of San Agustín National University (Cátedra Concytec);

e-mail: [jawitejada@hotmail.com](mailto:jawitejada@hotmail.com)

*Major Fields of Scientific Research: text mining, business intelligence, classification*