Detecting Errors in Foreign Trade Transactions: Dealing with Insufficient Data

Luis Torgo[1,2], Welma Pereira[1], and Carlos Soares[1,3]

[1] LIAAD-INESC Porto, Univ. of Porto, R. Ceuta, 118, 6., 4050-190 Porto, Portugal
[2] Faculdade de Ciências, University of Porto
[3] Faculdade de Economia, University of Porto
`ltorgo@liaad.up.pt, welma.pereira@gmail.com, csoares@fep.up.pt`
WWW home page: `http://www.liaad.up.pt/{~ltorgo,~csoares}`

Abstract. This paper describes a data mining approach to the problem of detecting erroneous foreign trade transactions in data collected by the Portuguese Institute of Statistics (INE). Erroneous transactions are a minority, but still they have an important impact on the official statistics produced by INE. Detecting these rare errors is a manual, time-consuming task, which is constrained by a limited amount of available resources (e.g. financial, human). These constraints are common to many other data analysis problems (e.g. fraud detection). Our previous work addresses this issue by producing a ranking of outlyingness that allows a better management of the available resources by allocating them to the most relevant cases. It is based on an adaptation of hierarchical clustering methods for outlier detection. However, the method cannot be applied to articles with a small number of transactions. In this paper, we complement the previous approach with some standard statistical methods for outlier detection for handling articles with few transactions. Our experiments clearly show its advantages in terms of the criteria outlined by INE for considering any method applicable to this business problem. The generality of the approach remains to be tested in other problems which share the same constraints (e.g. fraud detection).

1 Introduction

Portuguese companies have to declare their transactions with other EU countries on a monthly basis. The Portuguese Institute of Statistics (INE) collects data about the transactions of Portuguese companies with other EU countries on a monthly basis. These data typically contain errors, with various causes (e.g. typing errors). Due to the impact of these errors on official statistics, the data has to be manually inspected with the goal of trying to detect and correct them. This paper describes a data mining approach to help in this detection task. The objective is to identify the transactions that are most likely to contain errors. The selected transactions will then be manually analyzed by specialized staff and corrected if an error really exists.

Informally, the task is to detect transactions whose values are abnormal when compared to the other transactions of the same type of articles (products). This

task can be cast as an outlier detection problem. However, it has an important characteristic which significantly differentiates it from basic outlier detection problems: it is constrained by a limited amount of expensive human resources, whose availability varies at different moments.

Outlier detection is a well studied topic (e.g. [1]). Different approaches have been followed for this task. Distribution-based approaches (e.g. [2]) assume a certain parametric distribution of the data and signal outliers as observations that deviate from this distribution. The effectiveness of these approaches is highly dependent on how closely the distribution of the data follows the one assumed by the method (e.g. Gaussian). Knorr and Ng [3] introduced distance-based outlier detection methods. These approaches generalize several notions of distribution-based methods but still suffer from several problems, namely when the density of the data points varies (e.g. [4]). Breunig and colleagues [5, 4] defined density-based local outliers that handle this type of outliers. They are typically more appropriate than the previous ones for data with a complex distribution structure. The key idea of their work is that the notion of outlier should be "local" in the sense that the outlier status of any observation should be determined by the clustering structure in a bounded neighborhood of the observation.

Many of the existent outlier detection methods provide yes/no answers. For our application, this type of answers leads to sub-optimal decisions when it comes to manually inspecting the signaled cases. In effect, if the resources are limited we may well get more signals than what we can inspect and then we have to arbitrarily decide which cases to handle. In the work mentioned above [5, 4], the authors defined the notion of Local Outlier Factor (LOF) for each observation, which assigns a level of outlierness to each observation. This naturally leads to the notion of outlier ranking, which, in our problem, allows the resources to be used on the top rank cases. This type of constraints is not particular of this application. Similar setups can be found in many other application areas, namely on fraud detection. Previous work on this problem has compared outlier detection methods, a decision tree induction algorithm and a clustering method [6]. The results obtained with the latter did not satisfy the minimum goals, and, thus, the approach was dropped. Loureiro *et. al.* [7] have investigated more thoroughly the use of clustering methods to address this problem, achieving a significant boost of the results. Torgo [8] has proposed an improvement of the method described in [7] to obtain degrees of outlyingness. Torgo and Soares [9] have recently applied this later methodology to the INE problem with good results. However, all the methods proposed earlier cannot be applied to articles with a small number of transactions. In this paper we extend this work by proposing a new form of handling the articles with a small number of transactions.

2 Background

2.1 The INTRASTAT Application

Transactions made by Portuguese companies with organizations from other EU countries are declared monthly to INE using the INTRASTAT form. Using this

form, companies provide information about each transaction, like the article (item or product) ID, weight of the traded goods, total cost, etc.. At INE, these data are inserted into a database.

A problem with this process is that errors often occur in the process of filling forms. For instance, an incorrectly introduced item ID will associate a transaction with the wrong item. Therefore, when all of the transactions relative to a month have been entered into the database, they are manually verified with the aim of detecting and correcting as many errors as possible. In this search, the experts try to detect unusual values in the values that describe the transactions. Given that the number of transactions declared monthly is in the order of tens of thousands, this is a very costly process.

Besides kindly giving us 8 months of data, INE domain experts provided a few guidelines based on their experience with the problem. Namely,

- Among the different information provided by companies the cost per weight (Cost/Weight) item is the key attribute that they use to detect errors.
- Transactions should be analyzed by article, as different articles have significantly diverse ranges of acceptable values for Cost/Weight.
- If an article has less than 10 transactions all transactions are manually inspected.[4]

The general goal of this application is to use data mining for a better management of the resources involved in the inspection task. The system developed should be used monthly to automatically select a subset of the transactions for manual inspection. According to INE experts, to be acceptable such a system should select less than 50% of the transactions containing at least 90% of the errors. However, given that the available human resources are quite expensive and vary across different months, the smaller the number of selected transactions, the better, assuming that 90% of the errors are still detected. We will refer to the percentage of selected transactions by any method as %S, and the corresponding percentage of detected errors as the recall (%R) of the method. As we have 8 months of available data, this means that all candidate methods will be characterized by 8 pairs of %S and %R values. All values of %S should be below 50% and the corresponding values of %R above 90% to meet the operational requirements of INE.

One important issue concerning the evaluation of the results is that only a proportion of the transactions in the data provided have been manually inspected by INE experts. This means that if a transaction is flagged as an error, it is definitely an error. However, if the transaction is not flagged as an error it may have been inspected and found correct or it was simply not inspected and thus can actually be an error or not. In spite of this limitation of the available data, we can use this error information to evaluate the recall (%R) of any method. In effect, given a set of transactions that are selected by any method for manual inspection in a given month, we can calculate the recall as the percentage of the total number of flagged errors for that month that are contained on this set.

---

[4] Throughout the rest of the paper we will refer to these as the infrequent articles.

2.2 Previous Results

In [6] a first approach to this problem was described. The data used in that study contained transactions of five months in 1998. Four very different methods were applied. Two come from statistics and are univariate techniques: box plot [10] and Fisher's clustering algorithm [11]. The third one, Knorr & Ng's cell-based algorithm [3], is an outlier detection algorithm which, despite being a multivariate method, was used only on the Cost/Weight attribute. The last is C5.0 [12], a multivariate technique for the induction of decision trees. In this study, C5.0 obtained the best results, namely detecting 90% of the errors by inspecting 40% of the transactions. The latest approach was the only that provided a ranking of the transactions, and, thus, enabled the selection of the most suspicious transactions to be manually inspected depending on the available resources.

Loureiro *et. al* [7] have proposed a new outlier detection method based on the outcome of agglomerative hierarchical clustering methods and applied it to this problem. This approach used the size of the resulting clusters as indicators of the presence of outliers. The basic assumption was that outlier observations, being observations with unusual values, would be distant (in terms of the metric used for clustering) from the "normal" and more frequent observations, and therefore would be isolated in smaller clusters. These authors have explored several settings concerning the clustering process and experimentally evaluated them on the INTRASTAT problem. The best setup met the operational requirements of INE by detecting 94.1% of the errors on 32.7% of the transactions. In spite of this excellent result, the main drawback of this approach is that it simpy classifies each transaction as suspect or not. Thus, the set of selected transactions cannot be adapted to the inspection resources available. For instance, it may be the case that in a given month, there are not enough resources to inspect 32.7% of the transactions. In this case, we face the un-guided task of deciding which of these transactions will be inspected.

Torgo and Soares [9] have recently described a first approach to this resources constraint by applying an outlier ranking method [8], OF_H, to this problem. This method achieved competitive results when compared to a state of the art outlier ranking method like LOF ([4]), and satisfied the criteria established by INE. In this paper we try to improve these results even further.

3 Our Proposal

All previous approaches to this application have followed the guidelines given by the INE experts, in particular, that transactions of infrequent articles (i.e. with less than 10 transactions) are all manually inspected (Section 2.1). The reason for this recommendation is that it is difficult to automatically determine what is the acceptable range of values and, thus, what is an abnormal value.

However, the cost incurred by this strategy is very high. Table 1 presents the percentage of selected transactions (%S) in each of the 8 available months, by simply using this INE policy. Row %R reports the correspondent percentage

Table 1. The effort associated with selecting all transaction of infrequent articles.

	Jan/1998	Feb/1998	Mar/1998	May/1998	Jun/1998	Aug/1998	Sep/1998	Oct/1998
%S	35.7	30.8	27.7	24.5	32	21.0	17.0	22.5
%R	35.4	40.4	38.7	29.7	37	30.8	25.4	27.9
N_R	28	44	58	47	64	40	48	72

of errors detected by this strategy, while in row N_R we indicate the concrete number of errors that are found. Table 1 shows that this automatic selection strategy immediately imposes a large cost in terms of %S with results in terms of recall that are far from the target 90%. In effect, given that INE requires that the inspection effort should be less than 50% for a method to be acceptable, we see that in several months we already spend around 30% of this effort by simply including all transactions of infrequent articles. This means that there are lots of articles on these conditions, i.e. with less than 10 transactions per month. This motivates our decision to investigate approaches to replace this INE policy. Here, we propose methods to automatically process these infrequent articles and show that, in combination with our previous approach, they produce better results.

The main hypothesis is that it is possible to eliminate a significant amount of transactions of infrequent articles in an automatic and statistically grounded way. The resources that are saved in this way, can be used to inspect suspicious transactions that are detected in the remaining articles, which would, otherwise, not be inspected. The main difficulty we face is the fact that we have to make decisions based on very small amount of information. In effect, for these infrequent articles we have to decide whether a value is or not "suspicious" based on a sample of 10 or less values. Still, we claim that it must be possible to do better than just inspecting all of them.

We maintain the approach of handling the transactions of each article differently, depending on their number. If it is an infrequent article (less than 10 transactions) instead of automatically sending the transactions for manual inspection we propose some methods described in Section 3.1; if it is not an infrequent article we apply the outlier ranking method used in [9] that we describe in Section 3.2.

3.1 Handling Infrequent Articles

Methods for handling infrequent articles are constrained by the lack of data for these articles, which means that they must be simple. Therefore, we resort to simple univariate statistics heuristics. In this work, we used the box-plot rule that can be used to detect outliers in a continuous variable with a normal

distribution.[5] This rule declares a value as an outlier if it is outside the interval

$$[Q_1 - k \times IQR \cdots Q_3 + k \times IQR] \tag{1}$$

where Q_1 (Q_3) is the first (third) quartile of the variable, IQR is the interquartile range ($= Q_3 - Q_1$), and k is a constant (typically 1.5). The variable that we focus on is the Cost/Weight, which is the most informative according to the experts.

The small amounts of data reduces the power of the test (i.e., its ability to detect outliers).[6] This means that there is a compromise between finding errors in infrequent articles and reallocating resources to the remaining articles. Therefore, we have only applied the boxplot rule to infrequent articles with at least 5 transactions (i.e. values). If there are less than 5 transactions, they are all sent for inspection. Secondly, in order to try to avoid missing too many errors we have used a rather strict value of k, namely 0.0625 ($= 1.5/24$). This threshold will lead to selecting as potential outliers many of the values that would be ignored with more typical thresholds (e.g. 1.5), thus incurring a smaller risk of missing some of the flagged errors. In spite of these cautious settings we will see that this strategy has a strong impact on the number of transactions that are automatically selected for inspection. In our experiments we will call this the box plot (BP) method.

Additionally, we have created a variant of the BP method that decreases the number of infrequent articles. In the first month, it is equal to the BP method. In subsequent months, the data of infrequent articles is extended with data from previous months, referred to as *incremental dataset*. Our approach is then applied to the incremental dataset of each of those articles. This means that, if the number of transactions in the incremental dataset of a given infrequent article is 10 or more, it is handled as a frequent article. Otherwise, it continues to be handled as an infrequent article, as explained earlier.

This variant takes advantage of older data to generate more robust statistics, if the incremental dataset has less than 10 transactions. If the incremental dataset has 10 or more transactions, then it is possible to apply the more complex method for frequent articles.

We will call this the incremental box plot method (BP_{inc}). We should remark that in a real world application of this method we should limit the use of this past information. In effect, its justification lies on the assumption that the distribution of the "normal" values for the Cost/Weight attribute remains constant over time. While this is a reasonable assumption for most articles and for a period of 8 months, it may be inadequate for larger periods of time or for particular articles with temporary or permanent price volatility. In these cases a sliding window approach should probably be used instead.

3.2 Handling Frequent Articles

For articles with more data available we may use a more data-intensive approach to outlier detection. Clustering algorithms can be used to identify outliers as a

---

[5] This is a reasonable assumption for the cost per weight of the same article.
[6] If the number of values is less than 3, the method is not even applicable.

side effect of the clustering process. In this paper we use an approach that takes advantage of the dendrogram generated by hierarchical clustering methods to produce a ranking of outlyingness. This approach was first described in [8].

Agglomerative hierarchical clustering methods proceed in an iterative fashion by merging two of the current groups (which initially are formed by single observations) based on some criterion that is related to their proximity. This decision is taken locally, that is for each pair of groups, and takes into account the density of these two groups only. This merging process results in a tree-based structure usually known as a dendrogram. The merging step is guided by the information contained in the distance matrix of all available data. Several methods can be used to select the two groups to be merged at each stage. Contrary to other clustering approaches, hierarchical methods do not require a cluster initialization process that would inevitably spread the outliers across many different clusters thus probably leading to a rather unstable approach.

Informally, the idea behind this outlier detection approach is to use the height (in the dendrogram) at which any observation is merged into a group of observations as an indicator of its outlyingness. If an observation is really an outlier this should only occur at later stages of the merging process, that is the observation should be merged at a higher level than "normal" observations because it is very different from them and thus there is a "strong resistance" to joining their group. Formally, whenever the clustering algorithm decides to merge a group of observations g_s with another group g_l at iteration i of the merging process, where $|g_s| < |g_l|$, we set the outlier factor of the members of these groups as,

$$of_i(x) = \begin{cases} \left(1 - \frac{|g_s|}{N}\right) \times \frac{i}{N} & \text{if } x \in g_s \wedge |g_s| < t \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $|g_s|$ is the cardinality of the smallest group, g_s, t is a threshold that indicates the number of observations above which a group can not be regarded as a set of outliers for the data set, i is the step where the merge occurs, and N is the data set size.

Any observation can belong to several groups along its upward path through the dendrogram. As such, it will probably get several of these scores at different levels. These different values are in effect related to the issue of local and global outliers [8]. An observation that at a certain step gets a high score, may well have much smaller values up in the hierarchy. This happens frequently with local outliers. Until they are merged with their larger neighborhood they are regarded as outliers. After the merge, i.e. at a more global perspective, their diversity is diluted in the large cluster to which they now belong and thus they stop getting large scores. Still, these cases are outliers (local outliers), and thus they should get an overall large outlyingness score. In order to cope with these different types of outliers we set the outlyingness factor of any observation as the maximum $of_i()$ score it got along its path through the dendrogram. This means that the outlyingness factor of an observation is given by,

$$OF_H(x) = \max_i \ of_i(x) \tag{3}$$

Table 2. The results of the BP method.

		Jan/1998	Feb/1998	Mar/1998	May/1998	Jun/1998	Aug/1998	Sep/1998	Oct/1998
%S	INE	35.7	30.8	27.7	24.5	32	21.0	17.0	22.5
	BP	25.3	21.5	19.3	16.8	22.4	13.8	10.9	15.2
%R	INE	35.4	40.4	38.7	29.7	37	30.8	25.4	27.9
	BP	35.4	39.4	36.7	28.5	35.8	30.0	24.9	27.5
N_R	INE	28	44	58	47	64	40	48	72
	BP	28	43	55	45	62	39	47	71

One of the main advantages of this method is that we can use a standard hierarchical clustering algorithm to obtain the OF_H values without any additional computational cost. Here, we use a standard algorithm, with a time complexity of $O(N^2)$ and a space complexity of $O(N)$ [13]. We use the *hclust()* function of the statistical software environment R [14]. This function includes in its output a matrix (called **merge**) that can be used to easily obtain the necessary values for calculating directly the value of OF_H according to Equations 2 and 3.

4   Experimental Evaluation

In our experimental evaluation we address 3 main questions: i) the impact of our proposal to handle infrequent articles; ii) the overall performance of our error detection methodology when compared to state of the art competitors; and iii) the minimal effort that is required by our methods to attain the operational criteria required by INE (i.e. 90% recall).

We use the same experimental methodology as in previous studies with this data: we handle each month and article in turn, and we assess performance (i.e., the recall, %R, achieved for a certain inspection effort, %S, for each month), based on the errors identified by INE experts, as discussed earlier.

4.1   Different Approaches to Infrequent Articles

Our first experiments were designed to evaluate the impact of the different alternative approaches to handle infrequent articles that were described in Section 3.1.

Table 2 compares the default INE policy with the BP method. As you may confirm, this method achieves practically the same recall level, using considerably less inspection resources. This clearly indicates its advantage over the default INE policy and, moreover, it reduces the inspection effort dedicated to infrequent articles, thus enabling the inspection of more suspicious transactions from frequent articles.

Next, we evaluate the BP_{inc} method, which uses transactions from previous months to increase the amount of data available, as described earlier. For fairness, we compare BP_{INC} with the default policy of INE only on those articles that

Table 3. The results of the BP_{inc} method.

		Jan/1998	Feb/1998	Mar/1998	May/1998	Jun/1998	Aug/1998	Sep/1998	Oct/1998
%S	INE_{inc}	35.7	18.0	11.1	7.6	6.3	4.1	2.6	2.5
	BP_{inc}	25.3	11.8	6.8	4.8	3.9	2.2	1.3	1.5
%R	INE_{inc}	35.4	26.6	21.3	12.7	9.2	6.9	5.3	4.7
	BP_{inc}	35.4	22.9	18.7	11.4	8.1	6.9	5.3	3.9
N_R	INE_{inc}	28	29	32	20	16	9	10	12
	BP_{inc}	28	25	28	18	14	9	10	10

remain infrequent after being extended with data from previous months. Table 3 shows the results of this comparison. Again we observe a considerable decrease in the inspection effort with a small impact in terms of recall.

In summary, both our proposals show a clear advantage over the default INE policy of handling infrequent articles, by providing a similar recall with a significantly lower inspection effort.

4.2 Overall Comparison of Outlier Ranking Methods

In this section, we compare two outlier ranking (OR) methods to detect errors in frequent articles, in combination with the BP method on infrequent articles. The first method is the clustering-based method described earlier and the second is a state of the art competitor that was also used in the work by Torgo and Soares [9], LOF [4]. Namely, we have used the implementation of LOF that is available in R on package **dprep** [15].

Both OR methods were only applied on articles with more than 10 transactions. As they both produce rankings of the transactions, we can select the amount of inspection effort to use in accordance to our current resources constraints. Given INE requirements that this should not be above 50% of the transactions in each month, we have arbitrarily selected selected 4 effort levels to present our results: 35%, 40%, 45% and 50% of the transactions. For each of these %S levels we have calculated the corresponding recall scores (%R) achieved by each competitor. We have tried out 14 variants of their parameters and the results represent the best score obtained by each competitor. For infrequent articles we have used our BP method.

The results of this comparative experiments are shown in Figure 1. The two variants are denoted by OF_H and LOF, respectively. The four points for both alternatives represent the four previously mentioned working points in terms of %S. Still, we should recall that both methods would be better represented by lines as any other working points could have been selected. Some of the points are not shown on some graphs because the respective method achieved a very poor score that is outside of the used axes limits.

In the graphs of results for each month we also plot the %S and %R values of the method described in [7], which is denoted in the graphs as "LTS04". This

method is not an outlier ranking algorithm. It simply outputs the transactions it judges as being outliers, which leads to a pair of %S and %R values. In this case the user is not able to adjust the %S value to the available resources. Fortunately, in none of the testing months the 50% limit of selected transactions was surpassed but with this type of methods there is no such guarantee in general. Additionally, in those months where the available resources are not sufficient to analyze all the transactions selected, the experts must decide which ones to let aside. All graphs have two dotted lines indicating the experts requirements (at least 90% recall and at most 50% selected transactions).
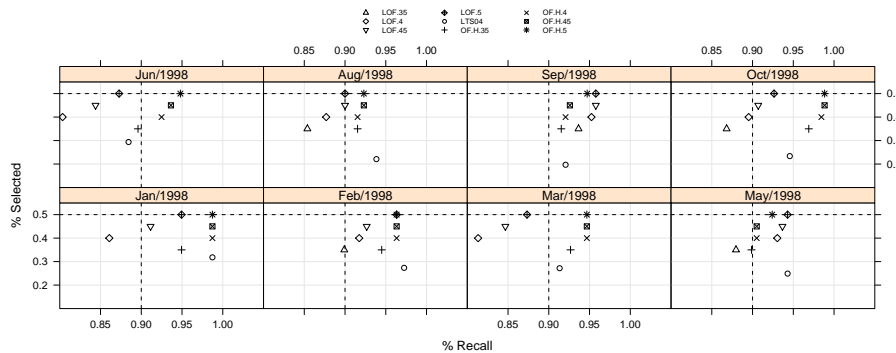

Fig. 1. The results of the comparative experiments on the INTRASTAT data.

The results of these experiments confirm that the clustering-based outlier ranking method is quite competitive with the state of the art. With the exception of two months (May and September) it consistently outperforms LOF. Moreover, it is the only method that always complies with INE operational criteria. Compared to the "LTS04" method [7], both OF_H and LOF loose a few times in terms of achieving the same %R for the same level of %S. Still, we should recall that "LTS04" provides no flexibility in terms of available human resources and thus it can happen (as for instance in Jun/1998) that the solution provided by this method does not satisfy the objectives of the experts or even that it is not feasible because it requires too many resources.

4.3   Minimum Required Effort for Satisfying the Recall Goal

In this section we address the issue of which is the minimum inspection effort that is required to achieve a value of 90% of recall. This score should always be below 50%, for any method to be applicable.

In these experiments we have considered both LOF and the OF_H method, in conjunction with the 3 alternative ways of handling the infrequent articles.

The results concerning this minimum effort levels are given in Figure 2. As we may confirm, independently of the infrequent articles method, the OF_H variants always achieve 90% of recall with less than 50% of effort. The same is not true for the LOF variants in March and June. This experiment also reveals the advantages of our methods for handling infrequent articles, which generally lead to better overall results. Namely, it is worth noting that with our methods in conjunction with OF_H, we achieve a recall of 90% with an effort around 30%, in general. With respect to the two alternative proposals (BP and BP_{inc}) the results are quite similar with no clear winner.
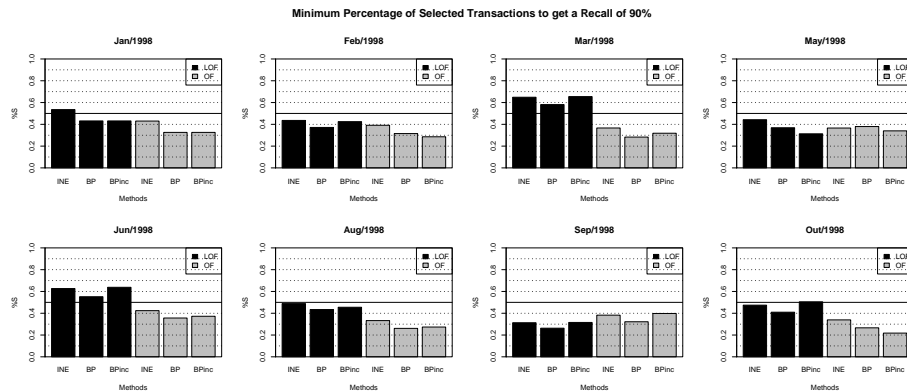

Fig. 2. Comparison between our method and LOF.

5 Conclusions

In many outlier and fraud detection problems there are situations in which there is plenty of data, but there are others where it is scarse (e.g., the number of credit card transactions of a new customer of a bank). In this paper, we extend previous work by proposing an approach that addresses the issue of outlier detection with scarse data. It involves the integration of simple statistical tests, used on small amounts of data, and a more complex outlier ranking method, used when sufficient data are available. The result is a method that can cope with variable resources for the necessary inspection phases. A few variations of the new approach were tested on the problem of detecting errors in foreign trade transaction forms of Portuguese companies. However, the approach is general and can be applied in different problems.

Compared to previous approaches to this problem, this work has introduced new forms of handling articles with few transactions. Our experiments have clearly shown their advantage when compared to the previous methods. The latter are based on domain knowledge provided by experts that all transactions

of infrequent articles should be analyzed. Our results contradict this, showing that it is possible to automate, at least partially, the processing of these articles.

Additionally, our experiments have shown the competitiveness of our clustering-based outlier ranking methodology when compared to other state of the art methods. Moreover, the proposal always complies with the operational requirements of INE and provides the desirable flexibility in terms of management of the available human inspection resources.

References

1. Victoria Hodge ; Jim Austin. A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22:85–126, 2004.
2. D. M. Hawkins. *Identification of Outliers*. Chapman and Hall, 11 New Fetter Lane, London EC4P 4EE, 1980.
3. Edwin M. Knorr and Raymond T. Ng. Algorithms for mining distance-based outliers in large datasets. In *Proceedings of 24rd International Conference on Very Large Data Bases (VLDB 1998)*, pages 392–403. Morgan Kaufmann, 1998.
4. M. M. Breunig, H. P. Kriegel, R. Ng, and J. Sander. Lof: Identifying density-based local outliers. In *Proceedings of ACM SIGMOD 2000 International Conference on Management of Data*, 2000.
5. M. M. Breunig, H. P. Kriegel, R. Ng, and J. Sander. Optics-of: Identifying local outliers. *Lecture Notes in Computer Science*, 1704:262–270, 1999.
6. C. Soares, P. Brazdil, J. Costa, V. Cortez, and A. Carvalho. Error detection in foreign trade data using statistical and machine learning methods. In N. Mackin, editor, *Proc. of the 3rd International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, pages 183–188, 1999.
7. A. Loureiro, L. Torgo, and C. Soares. Outlier detection using clustering methods: a data cleaning application. In Malerba D. and May M., editors, *Proceedings of KDNet Symposium on Knowledge-based Systems for the Public Sector*, 2004.
8. L. Torgo. Resource-bounded fraud detection. In Neves et. al, editor, *Proceedings of the 13th Portuguese Conference on Artificial Intelligence (EPIA'07)*, LNAI, pages 449–460. Springer, 2007.
9. L. Torgo and C. Soares. *Data Mining for Business Applications*, chapter Resource-bounded outlier detection using clustering methods. IOS Press, (to appear in 2009).
10. J.S. Milton, P.M. McTeer, and J.J. Corbet. *Introduction to Statistics*. McGraw-Hill, 1997.
11. W.D. Fisher. On grouping for maximum homogeneity. *Journal of the American Statistical Association*, 53:789–798, 1958.
12. R. Quinlan. *C5.0: An Informal Tutorial*. RuleQuest, 1998. http://www.rulequest.com/see5-unix.html.
13. F. Murtagh. Complexities of hierarchic clustering algorithms: state of the art. *Computational Statistics Quarterly*, 1:101–113, 1984.
14. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, 2008. ISBN 3-900051-07-0.
15. Edgar Acuna and members of the CASTLE group. *dprep: Data preprocessing and visualization functions for classification*, 2008. R package version 2.0.