Comparing Different Labeling Strategies in Anomalous Power Consumptions Detection

Fernanda Rodríguez$^{(\boxtimes)}$, Federico Lecumberry, and Alicia Fernández

Facultad de Ingeniería, Instituto de Ingeniería Eléctrica, Universidad de la República,
J. Herrera y Reissig 565, 11300 Montevideo, Uruguay
{fernandar,fefo,alicia}@fing.edu.uy

Abstract. Detecting anomalous events is a complex task, specially when it should be performed manually and for several hours. In the case of electrical power consumptions, the detection of non-technical losses also has a high economic impact. The diversity and big number of consumption records, makes it very important to find an efficient automatic method for detecting the largest number of frauds. This work analyses the performance of a strategy based on learning from expert labeling: suspect/no-suspect, with one using inspection labels: fraud/no-fraud. Results show that the proposed framework, suitable for imbalance problems, improves performance in terms of the $F_{measure}$ with inspection labels, avoiding hours of experts labeling.

Keywords: Electricity fraud · Support vector machine · Optimum Path Forest · Unbalance class problem · Combining classifier · UTE

1 Introduction

Non-technical loss detection is a huge challenge for electric power utility. In Uruguay the national electric company (henceforth UTE) faces the problem by manually monitoring a group of customers. A group of experts inspect at the monthly consumption curve of each customer and indicates those with some kind of suspicious behavior. This set of customers, initially classified as suspects are then analyzed taking into account other factors (such as fraud history, electrical energy meter type, etc.). Finally a subset of customers is selected to be inspected by an UTE's employee, who confirms (or not) the irregularity. The procedure is illustrated in Fig. 1. The procedure described before, has major drawbacks, mainly, the number of customers that can be manually controlled is small compared with the total amount of customer (around 500.000 only in Montevideo).

Different machine learning aproaches have addressed the detection of non-technical losses, both supervised or unsupervised. Leon et al. review the main research works found in the area between 1990 and 2008 [1]. Here we present a brief review that builds on this work and wide it with new contributions published between 2008 and 2013. Several of these approaches consider unsuper-vised classification using different techniques such as fuzzy clustering [2], neural
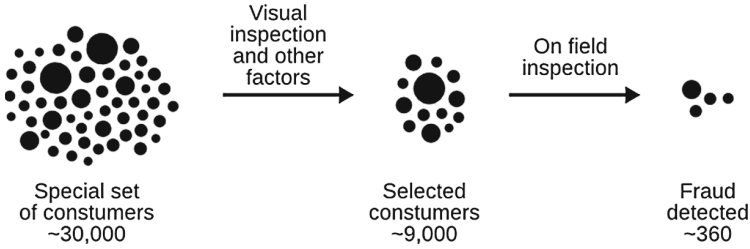
Fig. 1. Manual fraud detection scheme.

networks [3,4], among others. Monedero et al. use regression based on the correlation between time and monthly consumption, looking for significant drops in consumption [5]. Then they make a second stage where suspicious customers are eliminated if the consumption of these depend on the economy of the moment or the year's season. Only major customers were inspected and 38 % were detected as fraudulent. Similar results (40 %) were obtained in [6] using a tree classifier and customers who had been inspected in the past year. In [7,8] SVM is used. In the latter, Modified Genetic Algorithm is employed to find the best parameters of SVM. In [9], is compared the methods Back-Propagation Neural Network (BPNN), Online-sequential Extreme Learning Machine (OS-ELM) and SVM. Biscarri et al. [10] seek for outliers, Leon et al. [1] use Generalized Rule Induction and Di Martino et al. [11] combine CS-SVM classifiers, One class SVM, and C4.5 OPF using various features derived from the consumption. Different kinds of features are used among this works, for examples, consumption [8,10], contracted power and consumed ratio [12], Wavelet transformation of the monthly consumption [13], amount of inspections made to each client in one period and average power of the area where the customer resides [2], among others.

On the other hand, Romero proposes [14] a method to estimate and reduce non-technical losses, such as advanced metering infrastructure, fraud deterrence prepayment systems, system remote connection and disconnection, etc. Lo et al. based on real-time measurements, design [15] an algorithm for distributed state estimation in order to detect irregularities in consumption.

To improve the efficiency of fraud detection and resource utilization, in [16] was implemented a tool that automatically detects suspicious behavior analyzing customers historical consumption curve. This approach has the drawback of requiring a base previously tagged by the experts, in order to use it in the training stage.

In this work we set out to analyze the behavior of the proposed framework to fraud classification and compare it by using labels based on the inspection results instead of labels defined by experts. This new approach does not require that the company personnel conduct a manual study of the customers' consumption curve, since it use labels resulting from inspections in the past. We investigate performance improvement originated by training with individual algorithms and their combinations with labels of fraud and no fraud (based on inspections) and the importance of choosing the appropriate performance measure to solve the problem.

This paper is an extension of our previous work presented in the International Conference on Pattern Recognition Application and Methods (ICPRAM 2014) [17]; in Sect. 2 describes the framework and the strategies to be compare. Section 3 presents the obtained results and, finally Sect. 4 concludes the work.

2 Framework

The system presented consists basically on three modules: Pre-Processing and Normalization, Feature Extraction and Selection, and Classification. Figure 2 shows the system configuration. The system input corresponds to the last three years of the monthly consumption curve of each costumer.

The first module, Pre-Processing and Normalization, modifies the input data so that they all have normalized mean and implements some filters to avoid peaks from billing errors. A feature set was proposed taking into account UTE's technician expertize in fraud detection by manual inspection and recent papers on non technical loss detection [18–20]. Di Martino et al. use a list of the features extracted from the monthly consumption records [16]. In this work, we use the framework illustrated in Fig. 2 and a subset of the same set of features used in [16] but doing a selection of them taking into account the label type (based on inspection or expertise's criterion).

It is well known that finding a small set of relevant features can improve the final classification performance; this is why we implemented a feature selection stage. We used two types of evaluation methods: filter and wrapper. Filters methods looks for subsets of features with low correlation between them and high
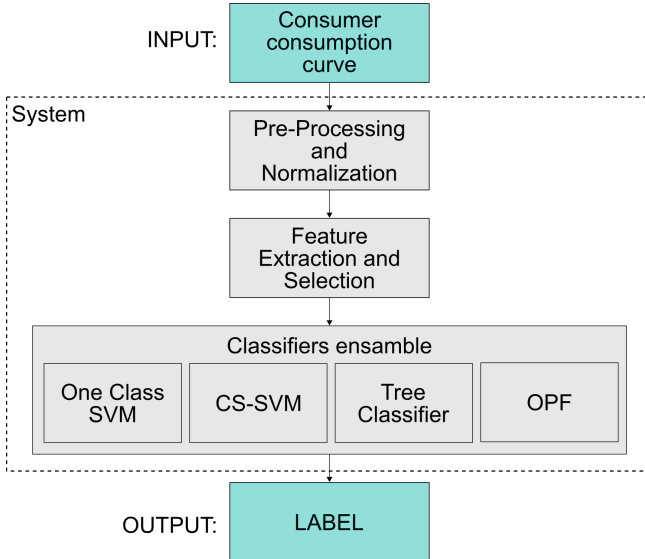

Fig. 2. Block diagram of the automatic fraud detection system.

correlation with the labels, while wrapper methods evaluate the performance of a given classifier for the given subset of features. In the wrapper methods, we used as performance measure the $F_{measure}$, also, the evaluations were performed using 10 fold cross validation over the training set.

As searching method, we used *Bestfirt*, for which we found in this application a good balance between performance and computational costs.

2.1 Classifiers

In this section we describe the classifiers used in this work. The authors of [21] proposed a new classifier, Optimum Path Forest (OPF), to apply to the problem of fraud detection in electricity consumption, showing good results. It consist of creating a graph with the training dataset, associating a cost to each path between two elements, based on the similarity between the elements of the path. This method assumes that the cost between elements of the same class is lower than those belonging to different classes. Next, a representative is chosen for each class, called prototypes. A new element is classified as the class that has lower cost with the corresponding prototype. Since OPF is very sensitive to class imbalance, we change class distribution of the training dataset by under-sampling the majority class.

The decision tree proposed by Ross Quinlan: C4.5 is used as another classifier. It is widely utilized because it is a very simply method that obtain good results. However, it is very unstable and highly dependent on the training set. Thus, a later stage of AdaBoost was implemented, accomplishing a more robust results. Just as with the previews classifier, it was needed an resamplig stage to manage the dependency of the C4.5 with the class distribution.

The other two classifiers consider the widely used method, SVM, cost-sensitive learning (CS-SVM) and one-class classifier (O-SVM). In the former different cost were assigned to the misclassification of the elements of each class, in order to tackle the unbalanced problem. The second one considers the minority class as the outliers.

Finally, we also consider another method, that performs an optimal combination of the before mentioned classifiers. Taking the labels of that each classifier find, the following functions are define:

$$g_p(x) = w_1^p d_{OPF}^p + w_2^p d_{Tree}^p + w_3^p d_{CS-SVM}^p + w_4^p d_{O_SVM}^p$$

$$g_n(x) = w_1^n d_{OPF}^n + w_2^n d_{Tree}^n + w_3^n d_{CS-SVM}^n + w_4^n d_{O_SVM}^n$$

where $d_i^j(x) = 1$ if the classifier j labels the sample as i and 0 otherwise. Then if $g_p(x) > g_n(x)$ the sample is assigned to the positive class, if $g_n(x) > g_p(x)$ the sample is assigned to the negative class. The choice of combination's weights (w_i^j) is done exhaustively in order to maximize the $F_{measure}$, over a predefined grid and was evaluated with a 10-fold cross validation.

2.2 The Class Imbalance Problem and the Choice of Performance Measure

When working on fraud detection problems, we can not assume that the number of people who commit fraud are near the same than those who do not, usually they are a minority class. This situation is known as class imbalance problem, and it is particularly important in real world applications where it is costly to misclassify examples from the minority class. In this cases, standard classifiers tend to be overwhelmed by the majority class and ignore the minority class, hence obtaining suboptimal classification performance. In order to confront this type of problem, different strategies can be used on different levels: (i) changing class distribution by resampling; (ii) manipulating classifiers; (iii) and on the ensemble of them, as proposed in [16].

Another problem which arises when working with imbalanced classes is that the most widely used metrics for measuring the performance of learning systems, such as *Accuracy* and *ErrorRate*, are not appropriate because they do not take into account misclassification costs, since they are strongly biased to favor the majority class [22]. Then others measures have to be considered:

– *Recall* is the percentage of correctly classified positive instances, in this case, the fraud samples.

$$Recall = \frac{TP}{TP + FN}$$

– *Precision* is defined as the proportion of labeled as positive instances that are actually positive.

$$Precision = \frac{TP}{TP + FP}$$

Where TP, FN and FP are defined in Table 1.
– The combination of this two measurements, the $F_{measure}$, represents the geometric mean between them, weighted by the parameter β,

$$F_{measure} = \frac{(1 + \beta^2)Recall \times Precision}{\beta^2\, Recall + Precision} \qquad (1)$$

Depending on the value of β we can prioritize *Recall* or *Precision*. For example, if we have few resources to perform inspections, it can be useful to prioritize *Precision*, so the set of samples labeled as positive has high density of true positive.

Table 1. Confusion matrix.

	Labeled as	
	Positive	Negative
Positive	TP (True Positive)	FN (False Negative)
Negative	FP (False Positive)	TN (True Negative)

When working with inspection labels the imbalance problem is worst, in terms of unbalance, than dealing with experts labels. In the experts labels method, the ratio of suspect to no suspect is near 10 %, while in the one based on inspection labels, the ratio is near 0.4 %.

3 Experiments and Results

In this work we used a data set of 456 industrial profiles obtained from the UTE's database. Each profile is represented by the customers monthly consumption in the last 36 months and has two labels, one dictated manually by technicians previous the inspection and another based on the inspection results. Training was done considering both labels separately and performance evaluation was done given the inspection labels, using a 10-fold cross validation scheme.
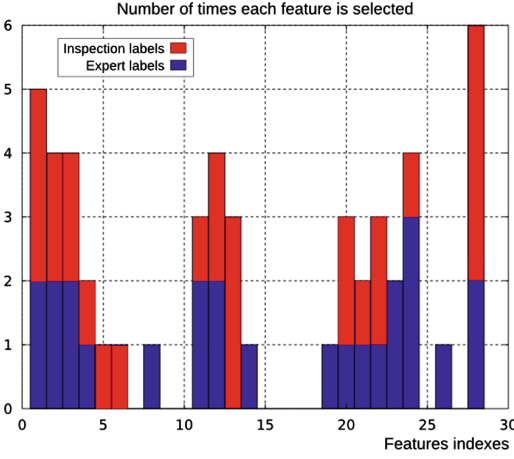
3.1 Features Selection

Different feature subsets were selected from the original set of 28 features proposed in [16], for the different classifiers and for both approaches. For example, for the experts' labels approach and the classifier CS-SVM, the features are:

– Consumption ratio for the last 3, 6 and 12 months and the average consumption (feature 1, 2 and 3).
– Difference between fourth Wavelet coefficient from the last and previous years (feature 11).
– Euclidean distance of each customer to the *mean customer*, where the *mean customer* is calculated by taking the mean for each month between all the customers (feature 20).
– Rate between the mean variance and the variance in the last year of the consumption curve (feature 21).
– Module of the first two Fourier coefficients (feature 23 and 24).
– Slope of the straight line that fits the consumption curve (feature 28).

While for inspection label approach and CS-SVM, the features are:

– Consumption ratio for the last 3 months and the average consumption (feature 1).
– Norm of the difference between the expected consumption and the actual consumption (feature 4).
– Difference between the third, fourth and fifth Wavelet coefficient from the last and previous years (feature 11, 12 and 13).
– Euclidean distance of each customer to the *mean customer* (feature 20).
– Ratio between the mean variance of each costumer and the mean variance of all the costumers, of the consumption curve (feature 22).
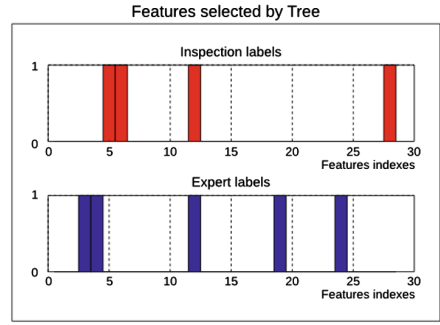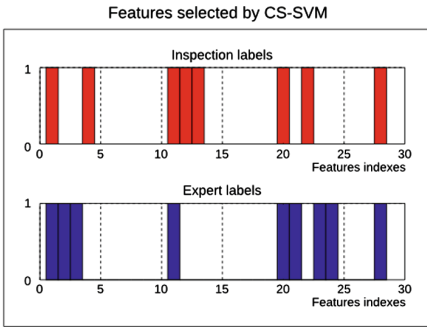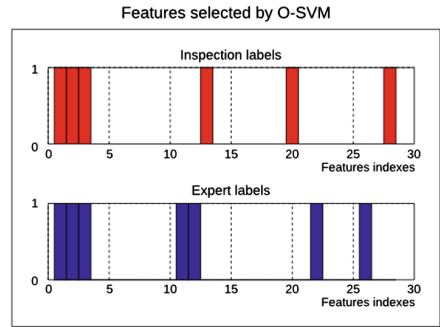– Slope of the straight line that fits the consumption curve (feature 28).
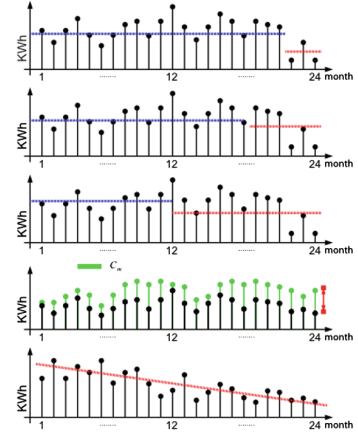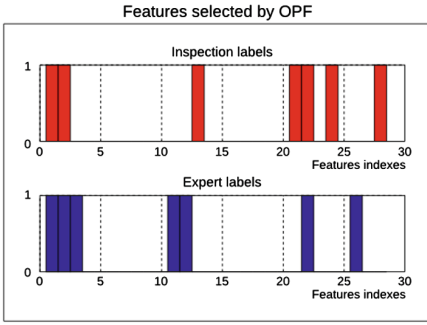
Fig. 3. Represents the features selection. (a) the abscissa indicates the 28 features and the ordinate is the number of times each feature is selected by the 4 classifiers and both methods (inspection label in red and expert label in blue). (b) outlines the features 1, 2, 3, and 28. (c), (d), (e) and (f) shows the features selected by each classifier and both methods separately (Color figure online).

Table 2. Fraud detection with experts label training.

Description	Recall (%)	Precision (%)	$F_{measure}$ (%) [$\beta = 1$]
OPF	39	27	32
Tree (C4.5)	38	23	29
O-SVM	51	22	30
CS-SVM	35	20	26
Iterative combination	77	22	35

Table 3. Fraud detection with inspection label training.

Description	Recall (%)	Precision (%)	$F_{measure}$ (%) [$\beta = 1$]
OPF	36	34	35
Tree (C4.5)	33	37	35
O-SVM	71	31	44
CS-SVM	74	33	46
Iterative combination	77	33	46

We observed that some features were selected for both approaches. Figure 3(d) represents the list indicated above, and Fig. 3(c), (d) and (f) indicates the features selected by OPF, Tree C.4.5 and O-SVM respectively, using the two different labels. Figure 3(a) shows the features selected by the four classifiers and both approaches all together. We can see that some features were never selected, other were selected only by one method and other by both method. Notice that each feature can be selected at the most eight times. The feature 28, represented by the fifth image in Fig. 3(b), is the most selected one (six times). Figure 3(b) also represents the feature 1, 2, 3 and 20. We conclude that, although the best subset of features depends on the classifiers and the type of label, some features are more representative than others.

3.2 Performance Analysis

Tables 2 and 3 shows the results obtained when experts and inspection labels are used to train the different classifiers respectively. The results for the method performed manually by experts, i.e. validating the expert labels with inspection labels, are $Recall = 38\%$, $Precision = 51\%$ and $F_{measure} = 44\%$.

In Table 2, we observe that the Iterative Combination technique with expert label training obtains the best result for fraud detection clearly overpassing the other methods, however the number false positive (FP) is relatively high, since

$$\frac{FP}{TP} = \frac{1}{Precision} - 1 \approx 4.$$

On the other hand, as Table 3 illustrates, if we use the inspection labels, the Iterative Combination also obtains the best results for fraud detection, but reducing in a half the number of FP ($\frac{FP}{TP} \approx 2$).

If we compare both approaches, we see that learning from the inspection labels could get better results (in the $F_{measure}$ sense) than learning from the labels set by experts. The former has the additional advantage of not requiring that the experts made the manual labeled of the training base.

Comparing the F_{masure} obtained manually by the experts (44%) and automatically by the Iterative Combination (46%) both are similar. However, the former consider other features as the history's fraud detection, contracted power, number of estimated readings, etc. and not only the monthly consumption, as the automatic one.

4 Conclusions and Future Work

In this work we compare the performance of a strategy based on learning from expert labeling: suspect/no-suspect, with one using inspection labels: fraud/no-fraud. In the $F_{measure}$ sense with all the tested classifiers the classification with inspection label obtains better results than using experts labels. Among them the Iterative Combination obtains the best result and also better than the manual method.

In future work we propose to include new categorical attributes as the history's fraud detection, contracted power, number of estimated readings, etc. We also want to explore a semi-supervised approach that allows to learn from data with and without previous inspection labels.

References

1. Leon, C., Biscarri, F.X.E.L., Monedero, I.X.F.I., Guerrero, J.I., Biscarri, J.X.F.S., Millan, R. X.E.O.: Variability and trend-based generalized rule induction model to NTL detection in power companies (2011)
2. dos Angelos, E., Saavedra, O., Corts, O., De Souza, A.: Detection and identification of abnormalities in customer consumptions in power distribution systems (2011)
3. Markoc, Z., Hlupic, N., Basch, D.: Detection of suspicious patterns of energy consumption using neural network trained by generated samples (2011)
4. Sforna, M.: Data mining in power company customer database. Electr. Power Syst. Res. **55**(3), 201–209 (2000)
5. Monedero, I., Biscarri, F., León, C., Guerrero, J.I., Biscarri, J., Millán, R.: Using regression analysis to identify patterns of non-technical losses on power utilities. In: Setchi, R., Jordanov, I., Howlett, R.J., Jain, L.C. (eds.) KES 2010, Part I. LNCS, vol. 6276, pp. 410–419. Springer, Heidelberg (2010)
6. Filho, J.R., Gontijo, E.M., Delaiba, A.C., Mazina, E., Cabral, J.E., Pinto, J.O.P.: Fraud identification in electricity company customers using decision tree (2004)
7. Depuru, S.S.S.R., Wang, L., Devabhaktuni, V.: Support vector machine based data classification for detection of electricity theft (2011)
8. Yap, K.S., Hussien, Z., Mohamad, A.: Abnormalities and fraud electric meter detection using hybrid support vector machine and genetic algorithm (2007)
9. Yap, K.S., Tiong, S.K., Nagi, J., Koh, J.S.P., Nagi, F.: Comparison of supervised learning techniques for non-technical loss detection in power utility (2012)
10. Biscarri, F., Monedero, I., Leon, C., Guerrero, J.I., Biscarri, J., Millan, R.: A data mining method based on the variability of the customer consumption - a special application on electric utility companies. In: Volume AIDSS, pp. 370–374. Inst. for Syst. and Technol. of Inf. Control and Commun. (2008)
11. Di Martino, J., Decia, F., Molinelli, J., Fernández, A.: Improving electric fraud detection using class imbalance strategies. In: 1st International Conference in Pattern Recognition Aplications and Methods, vol. 2, pp. 135–141 (2012)

12. Galvn, J., Elices, E., Noz, A.M., Czernichow, T., Sanz-Bobi, M.: System for detection of abnormalities and fraud in customer consumption (1998)
13. Jiang, R., Tagaris, H., Laschusz, A.: Wavelets based feature extraction and multiple classifiers for electricity fraud detection (2002)
14. Romero, J.: Improving the efficiency of power distribution system through technical and non-technical losses reduction (2012)
15. Lo, Y.L., Huang, S.C., Lu, C.N.: Non-technical loss detection using smart distribution network measurement data. In: 2012 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia), pp. 1–5 (2012)
16. Di Martino, M., Decia, F., Molinelli, J., Fernández, A.: A novel framework for nontechnical losses detection in electricity companies. In: Latorre Carmona, P., Sánchez, J.S., Fred, A.L.N. (eds.) Pattern Recognition - Applications and Methods. AISC, vol. 204, pp. 109–120. Springer, Heidelberg (2013)
17. Rodriguez, F., Lecumberry, F., Fernndez, A.: Non technical loses detection: experts labels vs. inspection labels in the learning stage (2014)
18. Alcetegaray, D., Kosut, J.: One class SVM para la detección de fraudes en el uso deenergía eléctrica. Trabajo Final Curso de Reconocimiento de Patrones, Dictado por el IIE- Facultad de Ingeniería- UdelaR (2008)
19. Muniz, C., Vellasco, M., Tanscheit, R., Figueiredo, K.: Ifsa-eusflat 2009 a neuro-fuzzy system for fraud detection in electricity distribution (2009)
20. Nagi, J., Mohamad, M.: Nontechnical loss detection for metered customers in power utility using support vector machines. IEEE Trans. Power Deliv. **25**(2), 1162–1171 (2010)
21. Ramos, C., de Sousa, A.N., Papa, J., Falcao, A.: A new approach for nontechnical losses detection based on optimum-path forest. IEEE Trans. Power Syst. (2010)
22. Garcia, V., Sanchez, J., Mollineda, R.: On the suitability if numerical performance evaluation measures for class imbalance problems. In: 1st International Conference in Pattern Recognition Aplications and Methods, vol. 2, pp. 310–313 (2012)