The Challenge of Non-Technical Loss Detection using Artificial Intelligence: A Survey

Patrick Glauner, *Member, IEEE,* Andre Boechat, Lautaro Dolberg, Jorge Meira, Radu State, Franck Bettinger, Yves Rangoni, and Diogo Duarte

***Abstract*—Detection of non-technical losses (NTL) which include electricity theft, faulty meters or billing errors has attracted increasing attention from researchers in electrical engineering and computer science. NTLs cause significant harm to the economy, as in some countries they may range up to 40% of the total electricity distributed. The predominant research direction is employing artificial intelligence (AI) to solve this problem. Promising approaches have been reported falling into two categories: expert systems incorporating hand-crafted expert knowledge or machine learning, also called pattern recognition or data mining, which learns fraudulent consumption patterns from examples without being explicitly programmed. This paper first provides an overview about how NTLs are defined and their impact on economies. Next, it covers the fundamental pillars of AI relevant to this domain. It then surveys these research efforts in a comprehensive review of algorithms, features and data sets used. It finally identifies the key scientific and engineering challenges in NTL detection and suggests how they could be solved. We believe that those challenges have not sufficiently been addressed in past contributions and that covering those is necessary in order to advance NTL detection.**

***Index Terms*—Artificial intelligence, data mining, electricity theft, expert systems, machine learning, non-technical losses, pattern recognition.**

I. Introduction

OUR modern society and daily activities strongly depend on the availability of electricity. Electrical power grids allow to distribute and deliver electricity from generation infrastructure such as power plants or solar cells to customers such as residences or factories. One frequently appearing problem are losses in power grids, which can be classified into two categories: technical and non-technical losses.

Technical losses occur mostly due to power dissipation. This is naturally caused by internal electrical resistance and the affected components include generators, transformers and transmission lines.

The opposite class of losses are non-technical losses (NTL), which are primarily caused by electricity theft. In most countries, NTLs account for the predominant part of the overall losses [46]. Therefore, it is most beneficial to first reduce NTLs before reducing technical losses [3]. Nonetheless, reducing technical losses is challenging, too. In particular, NTLs include, but are not limited to, the following causes [15], [65]:

- Meter tampering in order to record lower consumptions
- Bypassing meters by rigging lines from the power source
- Arranged false meter readings by bribing meter readers
- Faulty or broken meters
- Un-metered supply
- Technical and human errors in meter readings, data processing and billing

NTLs cause significant harm to economies, including loss of revenue and profit of electricity providers, decrease of the stability and reliability of electrical power grids and extra use of limited natural resources which in turn increases pollution. There are different estimates of the losses caused by NTL. For example, in India, NTLs are estimated at US$ 4.5 billion [8]. NTLs also reported to range up to 40% of the total electricity distributed in countries such as Brazil, India, Malaysia or Lebanon [22], [45]. They are also of relevance in developed countries, for example estimates of NTLs in the UK and US range from US$ 1-6 billion [2], [46].

In order to detect NTLs, inspections of customers are carried out based on predictions. From an electrical engineering perspective, one method to detect losses is to calculate the energy balance [58], which requires topological information of the network. This does not work accurately for those reasons: (i) network topology undergoes continuous changes in order to satisfy the rapidly growing demand of electricity, (ii) infrastructure may break and lead to wrong energy balance calculations and (iii) it requires transformers, feeders and connected meters to be read at the same time. A more flexible and adaptable approach is to employ artificial intelligence (AI) [62]. AI allows to analyze customer profiles, their data and known irregular behavior in order to trigger a possible inspection of a customer. However, carrying out inspections is costly, as it requires physical presence of technicians. It is therefore important to make accurate predictions in order to reduce the number of false positives.

The rest of this paper is organized as follows. Section II describes the field of AI. Section III provides a detailed review and critique of state-of-the-art NTL detection research employing AI methods. In Section IV, we identify the key challenges of this field that need to be accurately studied in order to enhance methods in the future. To the best of our knowledge, this topic has not been addressed yet in the literature on NTL detection. Section V summarizes this survey.

II. Artificial Intelligence

The field of artificial intelligence (AI) attempts to both understand and build intelligent entities [62]. This name was

P. Glauner, A. Boechat, L. Dolberg, J. Meira and R. State are with the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, (email: {first.last}@uni.lu).

F. Bettinger, Y. Rangoni and D. Duarte are with CHOICE Technologies Holding Sàrl, (email: {first.last}@choiceholding.com).

Manuscript received MONTH DD, YEAR.

coined in 1955 during the preparations for the first AI conference hosted at Darmouth College [39]. While most people intuitively think about robotics, AI has more applications, such as learning patterns from data. This chapter provides an overview of the AI methods relevant to NTL detection.

A. Expert systems

Traditional AI systems were based on hand-crafted rules. Such systems are also called expert systems because they incorporate expert knowledge in their decision making process. While expert systems have initially been successful in tasks such as diagnosis and treatment of nuclear reactor accidents [48] or mission planning of autonomous underwater vehicles [33], they have the following shortcomings: (1) incorporating expert knowledge in rules is challenging, (2) many domains cannot accurately be described in rules and (3) domain knowledge may change over time requiring amendments of the rules [31]. Nonetheless, expert systems are still being used nowadays.

B. Machine learning

To avoid the shortcomings of expert systems, a diametrically opposed approach is to learn patterns from data rather than hand-crafting rules. This branch of AI is called machine learning or pattern recognition. Both approaches have their justification and neither is generally better or worse than the other in artificial intelligence [25]. Machine learning gives computers the ability to learn from data without being explicitly programmed [50]. This property has allowed to significantly improve AI in various applications, such as in handwritten digit recognition [35], facial expression recognition [71] or speech recognition [27]. Machine learning consists of three pillars: supervised, unsupervised and reinforcement learning. The term data mining is strongly related to machine learning, but has a wider scope that includes data cleaning, data preprocessing and concrete applications.

1) Supervised learning: Supervised learning algorithms learn patterns from labeled training examples $(x^{(i)}, y^{(i)})$, in which $x^{(i)}$ is a training data point and $y^{(i)}$ is a corresponding label. This is also called function induction and typical applications include regression or classification [10]. This pillar is best understood at present time and there are a wide variety of available learning algorithms. The choice of which learning algorithm to apply to a concrete problem is challenging and often requires comparative experiments. However, having a lot of representative data is considered sometimes to be more relevant than the actual algorithm [7].

2) Unsupervised learning: Unsupervised learning uses only unlabeled data points $x^{(i)}$ in order to find hidden structure in the data [10]. Applications include dimensionality reduction methods such as the Principal Component Analysis (PCA) or t-sne [36] and clustering algorithms such as K-means.

3) Reinforcement learning: In many learning problems, there is no intuitively correct supervision. Reinforcement learning is a reward-based learning technique for actions in order to get to a goal [67]. It has for example successfully been applied to humanoid robotics [61], autonomous helicopter flying [51] and playing the game of Go at super-human performance [64].

III. The State of the Art

NTL detection can be treated as a special case of fraud detection, for which a general survey is provided in [11] and [32]. It highlights expert systems and machine learning as key methods to detect fraudulent behavior in applications such as credit card fraud, computer intrusion and telecommunications fraud. This section is focused on an overview of the existing AI methods for detecting NTLs. For other surveys of the past efforts in the field, readers are referred to [15] and [30]. Overviews of possible methods to manipulate a smart metering infrastructure are provided in [29] and [40].

A. Support Vector Machines

Support Vector Machines (SVM) [69] are a state-of-the-art classification algorithm that is less prone to overfitting. Electricity customer consumption data of less than 400 highly imbalanced out of ~260K customers in Kuala Lumpur, Malaysia having each 25 monthly meter readings in the period from June 2006 to June 2008 are used in [43]. From these meter readings, daily average consumptions features per month are computed. Those features are then normalized and used for training in a SVM with a Gaussian kernel. For this setting, a recall of 0.53 is achieved on the test set. In addition, credit worthiness ranking (CWR) is used in [46]. It is computed from the electricity provider's billing system and reflects if a customer delays or avoids payments of bills. CWR ranges from 0 to 5 where 5 represents the maximum score. It was observed that CWR significantly contributes towards customers committing electricity theft. A test accuracy of 0.77 and a test recall of 0.64 are reported.

SVMs are also applied on 1,350 Indian customer profiles in [21]. They are split into 135 different daily average consumption patterns, each having 10 customers. For each customer, meters are read every 15 minutes. A test accuracy of 0.984 is reported. This work is extended in [20] by encoding the $4 \times 24 = 96$-dimensional input in a lower dimension indicating possible irregularities. This encoding technique results in a simpler model that is faster to train while not losing the expressiveness of the data and results in a test accuracy of 0.92. This work is extended in [22] by introducing high performance computing algorithms in order to enhance the performance of the previously developed algorithms in [20]. This faster model has a test accuracy of 0.89.

Consumption profiles of 5K Brazilian industrial customer profiles are analyzed in [57]. Each customer profile contains 10 features including the demand billed, maximum demand, installed power, etc. In this setting, a SVM slightly outperforms K-nearest neighbors (KNN) and a neural network, for which test accuracies of 0.9628, 0.9620 and 0.9448, respectively, are reported.

B. Neural networks

Neural networks [9] are loosely inspired by how the human brain works and allow to learn complex hypotheses from

data. An ensemble of five neural networks (NN) is trained on samples of a data set containing ~20K customers in [41]. Each neural network uses features calculated from the consumption time series plus customer-specific pre-computed attributes. A precision of 0.626 and an accuracy of 0.686 are obtained on the test set.

A data set of ~22K customers is used in [17] for training a neural network. It uses the average consumption of the previous 12 months and other customer features such as location, type of customer, voltage and whether there are meter reading notes during that period. On the test set, an accuracy of 0.8717, a precision of 0.6503 and a recall of 0.2947 are reported.

Extreme learning machines (ELM) are one-hidden layer neural networks in which the weights from the inputs to the hidden layer are randomly set and never updated. Only the weights from the hidden to output layer are learned. The ELM algorithm is applied to NTL detection in meter readings of 30 minutes in [52], for which a test accuracy of 0.5461 is reported.

A self-organizing map (SOM) is a type of unsupervised neural network training algorithm that is used for clustering. SOMs are applied to weekly customer data of 2K customers consisting of meter readings of 15 minutes in [13]. This allows to cluster customers' behavior into fraud or non-fraud. Inspections are only carried out if certain hand-crafted criteria are satisfied including how well a week fits into a cluster and if no contractual changes of the customer have taken place. A test accuracy of 0.9267, a test precision of 0.8526, and test recall of 0.9779 are reported.

C. Expert systems and fuzzy systems

Profiles of 80K low-voltage and 6K high-voltage customers in Malaysia having meter readings every 30 minutes over a period of 30 days are used in [47] for electricity theft and abnormality detection. A test recall of 0.55 is reported. This work is related to features of [45], however, it uses entirely fuzzy logic incorporating human expert knowledge for detection.

A database of ~700K Brazilian customers, ~31M monthly meter readings from January 2011 to January 2015 and ~400K inspection data is used in [24]. It employs an industrial Boolean expert system, its fuzzified extension and optimizes the fuzzy system parameters using stochastic gradient descent [6] to that database. This fuzzy system outperforms the Boolean system. Inspired by [43], a SVM using daily average consumption features of the last 12 months performs better than the expert systems, too. The three algorithms are compared to each other on samples of varying fraud proportion containing ~100K customers. It uses the area under the (receiver operating characteristic) curve (AUC), which is discussed in Chapter IV-A. For a NTL proportion of 5%, it reports AUC test scores of 0.465, 0.55 and 0.55 for the Boolean system, optimized fuzzy system and SVM, respectively. For a NTL proportion of 20%, it reports AUC test scores of 0.475, 0.545 and 0.55 for the Boolean system, optimized fuzzy system and SVM, respectively.

Five features of customers' consumption of the previous six months are derived in [4]: average consumption, maximum consumption, standard deviation, number of inspections and the average consumption of the residential area. These features are then used in a fuzzy c-means clustering algorithm to group the customers into c classes. Subsequently, the fuzzy membership values are used to classify customers into NTL and non-NTL using the Euclidean distance measure. On the test set, an average precision (called average assertiveness) of 0.745 is reported.

The database of [41] is used in [42]. In the first step, an ensemble pre-filters the customers to select irregular and regular customers for training which represent well two different classes. This is done because of noise in the inspection labels. In the classification step, a neuro-fuzzy hierarchical system is used. In this setting, a neural network is used to optimize the fuzzy membership parameters, which is a different approach to the stochastic gradient descent method used in [24]. A precision of 0.512 and an accuracy of 0.682 on the test set are obtained.

The work in [46] is combined with a fuzzy logic expert system postprocessing the output of the SVM in [45] for ~100K customers. The motivation of that work is to integrate human expert knowledge into the decision making process in order to identify fraudulent behavior. A test recall of 0.72 is reported.

D. Genetic algorithms

The work in [43] and [46] is extended by using a genetic SVM in [44] for 1,171 customers. It uses a genetic algorithm in order to globally optimize the hyperparameters of the SVM. Each chromosome contains the Lagrangian multipliers $(\alpha_1, ..., \alpha_i)$, regularization factor C and Gaussian kernel parameter γ. This model achieves a test recall of 0.62.

A data set of ~1.1M customers is used in [18]. The paper identifies the much smaller sample of inspected customers as the main challenge NTL detection. It then proposes stratified sampling in order to increase the number of inspections and to minimize the statistical variance between them. The stratified sampling procedure is defined as a non-linear restricted optimization problem of minimizing the overall energy loss due to electricity theft. This minimization problem is solved using two methods: (1) genetic algorithm and (2) simulated annealing. The first approach outperforms the other one. Only the reduced variance is reported, which is not comparable to the other research and therefore left out of this survey.

E. Other methods

Optimum path forests (OPF), a graph-based classifier, is used in [54]. It builds a graph in the feature space and uses so-called "prototypes" or training samples. Those become roots of their optimum-path tree node. Each graph node is classified based on its most strongly connected prototype. This approach is fundamentally different to most other learning algorithms such as SVMs or neural networks which learn hyperplanes. Optimum path forests do not learn parameters, thus making training faster, but predicting slower compared to parametric

methods. They are used in [55] for 736 customers and achieved a test accuracy of 0.9021, outperforming SMVs with Gaussian and linear kernels and a neural network which achieved test accuracies of 0.8893, 0.4540 and 0.5301, respectively. Related results and differences between these classifiers are reported in [56].

Rough sets give lower and upper approximations of an original conventional or crisp set. Rough set analysis is applied to NTL detection in [66] on features related to [17]. This supervised learning technique allows to approximate concepts that describe fraud and regular use. A test accuracy of 0.9322 is reported. The first application of rough set analysis applied to NTL detection is described in [12] on 40K customers, but lacks details on the attributes used per customer, for which a test accuracy of 0.2 is achieved.

Different feature selection techniques for customer master data and consumption data are assessed in [53]. Those methods include complete search, best-first search, genetic search and greedy search algorithms for the master data. Other features called shape factors are derived from the consumption time series including the impact of lunch times, nights and weekends on the consumption. These features are used in K-means for clustering similar consumption time series. In the classification step, a decision tree is used to predict whether a customer causes NTLs or not. An overall test accuracy of 0.9997 is reported.

A different method is to estimate NTLs by subtracting an estimate of the technical losses from the overall losses [63]. It models the resistance of the infrastructure in a temperature-dependent model using regression which approximates the technical losses. It applies the model to a database of 30 customers for which the consumption was recorded for six days with meter readings every 30 minutes for theft levels of 1, 2, 3, 4, 6, 8 and 10%. The respective test recalls in linear circuits are 0.2211, 0.7789, 0.9789, 1, 1, 1 and 1, respectively.

F. Summary

A summary and comparison of the performance measures of selected classifiers discussed in this review are reported in Table I. The most commonly used models comprise Boolean and fuzzy expert systems, SVMs and neural networks. In addition, genetic methods, OPF and regression methods are used. Data set sizes have a wide range from 30 up to 700K customers. However, the largest data set of 1.1M customers in [18] is not included in the table because only the variance is reduced and no other performance measure is provided. Accuracy and recall are the most popular performance measures in the literature, ranging from 0.45 to 0.99 and from 0.29 to 1, respectively. Only very few publications report the recall of their models, ranging from 0.51 to 0.85. The AUC is only reported in one publication. The challenges of finding representative performance measures and how to compare individual contributions are discussed in Chapters IV-A and IV-F, respectively.

TABLE I
SUMMARY OF PERFORMANCE MEASURES (TWO-DECIMAL PRECISION)

Ref	Model	#Cust[a]	Acc[b]	Pre[c]	Rec[d]	AUC[e]	Prop[f]
[13]	SOM	2K	0.93	0.85	0.98	-	-
[17]	NN	22K	0.87	0.65	0.29	-	-
[21]	SVM (Gauss)	1,350	0.98	-	-	-	-
[24]	Bool rules	700K	-	-	-	0.47	5%
[24]	Fuzzy rules	700K	-	-	-	0.55	5%
[24]	SVM (linear)	700K	-	-	-	0.55	5%
[24]	Bool rules	700K	-	-	-	0.48	20%
[24]	Fuzzy rules	700K	-	-	-	0.55	20%
[24]	SVM (linear)	700K	-	-	-	0.55	20%
[42]	Neuro-fuzzy	20K	0.68	0.51	-	-	-
[43]	SVM	< 400	-	-	0.53	-	-
[44]	Genetic SVM	1,171	-	-	0.62	-	-
[45]	SVM + fuzzy	100K	-	-	0.72	-	-
[46]	SVM (Gauss)	< 400	0.77	-	0.64	-	-
[53]	Decision tree	N/A	0.99	-	-	-	-
[55]	OPF	736	0.90	-	-	-	-
[55]	SVM (Gauss)	736	0.89	-	-	-	-
[55]	SVM (linear)	736	0.45	-	-	-	-
[55]	NN	736	0.53	-	-	-	-
[57]	SVM	5K	0.96	-	-	-	-
[57]	KNN	5K	0.96	-	-	-	-
[57]	NN	5K	0.94	-	-	-	-
[63]	Regression	30	-	-	0.22	-	1%
[63]	Regression	30	-	-	0.78	-	2%
[63]	Regression	30	-	-	0.98	-	3%
[63]	Regression	30	-	-	1	-	4-10%
[66]	Rough sets	N/A	0.93	-	-	-	-

[a]Number of customers in the database used
[b]Accuracy
[c]Precision
[d]Recall
[e]Area under the receiver operating characteristic curve
[f]NTL/theft proportion

IV. CHALLENGES

The research reviewed in the previous section indicate multiple open challenges. These challenges do not apply to single contributions, rather they spread across different contributions. In this section, we focus on discussing these challenges, which are necessary in order to advance in NTL detection. Concretely, we discuss common topics that have not yet received the necessary attention in previous research and put them in the context of machine learning research as a whole.

A. Class imbalance and evaluation metric

Imbalanced classes appear frequently in machine learning, however, this fact is mostly not addressed in the literature. This topic is well covered for example in [28] and [68]. The class imbalance also affects the choice of evaluation metrics. Most NTL detection research such as [17], [18], [43], [54] and [66] also ignore this topic and report high accuracies or recalls. The following examples demonstrate why those performance measures are not suitable for NTL detection in imbalanced data sets: for a test set containing 1K customers of which 999 have regular use, (1) a classifier always predicting non-

NTL has an accuracy of 99.9%, whereas in contrast, (2) a classifier always predicting NTL has a recall of 100%. While the classifier of the first example has a very high accuracy and intuitively seems to perform very well, it will never predict any NTL. In contrast, the classifier of the second example will find all NTL, but trigger many costly and unnecessary physical inspections. This topic is addressed for example in [37] and [41], but do not use a proper single performance measure to describe the performance of a classifier performed on an imbalanced dataset.

For NTL detection, the goal is to reduce the false positive rate (FPR) to decrease the number of costly inspections, while increasing the true positive rate (TPR) to find as many NTL occurrences as possible. [24] proposes to use a receiver operating characteristic (ROC) curve, which plots the TPR against the FPR. The area under the curve (AUC) is a performance measure between 0 and 1, where any binary classifier with an AUC > 0.5 performs better than random guessing. In order to assess a NTL prediction model using a single performance measure, the AUC was picked as the most suitable one. In the preliminary work of [24], we noticed that the precision usually grows linearly with the NTL proportion in the data set. It is therefore not suitable for low NTL proportions. However, we did not notice this for the recall and made observations of non-linearity similar to the work of [63] summarized in Table I. With the limitation of precision and recall as isolated performance measures, the F_1 score did not prove to work as a reliable performance measure. We believe that it is necessary to investigate more into this topic in order to report reliable and imbalance-independent results that are valid for different levels of imbalance. The Matthews correlation coefficient (MCC) defined in [38]:

$$\frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (1)$$

measures the accuracy of binary classifiers taking into account the imbalance of both classes, ranging from -1 to $+1$. We believe that this measure should be assessed further for NTL detection.

B. Feature description

Different feature description methods have been reviewed in the previous section. Hand-crafting features from raw data is a long-standing issue in machine learning having significant impact on the performance of a NTL classifier [23]. Generally, it cannot easily be said if a feature description is good or bad. Deep learning allows to self-learn hidden correlations and increasingly more complex feature hierarchies from the raw data input [34]. This approach has lead to breakthroughs in image analysis and speech recognition [27]. One possible method to overcome the challenge of feature description for NTL detection is by finding a way to apply deep learning to it.

C. Incorrect inspection results

In the preliminary work of [24] we noticed that the inspection result labels in the training set are not always correct and that some fraudsters may be labelled as non-fraudulent. The reasons for this may include bribing, blackmailing or threatening of the technician performing the inspection. Also, the fraud may be done too well and it is not observable by technicians. Another reason may be incorrect processing of the data. It must be noted that the latter reason may, however, also label non-fraudulent behavior as fraudulent. Handling noise is a common challenge in machine learning. In supervised machine learning settings, most existing methods address handling noise in the input data. There are different regularization methods such as L_1 or L_2 regularization [49] or learning of invariances allowing learning algorithms to better handle noise in the input data [10], [35]. However, handling noise in the training labels is less commonly addressed in the machine learning literature. Most NTL detection research use supervised methods and this shortcoming of the training data and potential false positive labels in particular are not much reported in the literature, except in [42].

Unsupervised methods such as clustering or dimensionality reduction [36] that totally ignore the labels can be used to overcome this limitation. Also, in many situations, most customers have never been inspected. In a purely supervised training strategy, the unlabeled data is discarded. However, using it may support training. This domain is called semi-supervised, for which semi-supervised clustering [5] has been proposed. Deep neural networks can be pre-trained using autoencoders or restricted Boltzmann machines [34] in order to take advantage of unlabeled data. Both, unsupervised and semi-supervised learning, should be further explored for NTL detection. Furthermore, we are not aware of research that has applied reinforcement learning to NTL detection. We believe that it can be promising to explore in this direction, as reinforcement learning needs only very little supervision in the form of rewards.

D. Biased inspection results

In statistics, samples of data must represent the overall population in order to make valid conclusions. This is a long-standing issue in statistics and therefore in machine learning, too, as discussed in [26]. In the preliminary work of [24] we noticed that the sample of previously inspected customers may not be representative of all customers. One reason is, for example, that electricity suppliers previously focused on certain neighborhoods for inspections. NTL classifiers trained on biased inspection data are likely to be biased, too. To the best of our knowledge, this topic has not been addressed in the literature on NTL detection. Bias correction has initially been addressed in the field of computational learning theory [16]. We believe that it can be promising to explore in this direction. For example, one promising approach may be resampling inspection data in order to be representative in terms of location and type of customer.

E. Scalability

The number of customers used throughout the research reviewed significantly varies. For example, [43] and [63] only use less than a few hundred customers in the training. A SVM

with a Gaussian kernel is used in [43], for which training is only feasible in a realistic amount of time for up to a couple of ten thousand customers in current implementations [14]. A regression model using the Moore-Penrose pseudoinverse [59] is used in [63]. This model is also only able to scale for up to a couple of ten thousand customers [50]. Models being trained on up to a couple of ten thousand customers include [41] and [17] using neural networks. The training methods used in those papers usually do not scale to significantly larger customer databases. Larger databases using up to hundreds of thousand or millions of customers are used in [18] and [24] using a SVM with linear Kernel or genetic algorithms, respectively.

We believe that a stronger investigation into time complexity of learning algorithms, scalable computing models and technologies such as Apache Spark [70] or Google TensorFlow [1] will allow to efficiently handle Big Data sets for NTL detection. This will also allow to perform the computations in a cloud, requiring researchers to make significantly lower investments in hardware.

F. Comparison of different methods

Comparing the different methods reviewed in this paper is challenging because they are tested on different data sets, as summarized in Table I. In many cases, the description of the data lacks fundamental properties such as the number of meter readings per customer, NTL proportion, etc. In order to make results better comparable, joint efforts of different research groups are necessary in order to address the comparability of NTL detection system performance based on a comprehensive freely available and sufficiently large data set.

V. Conclusion

Non-technical losses (NTL) are the predominant type of losses in electricity power grids. We have reviewed their impact on economies and potential losses of revenue and profit for electricity providers. In the literature, a vast variety of NTL detection methods employing artificial intelligence methods are reported. Expert systems and fuzzy systems are traditional detection models. Over the past years, machine learning methods have become more popular. The most commonly used methods are support vector machines and neural networks, which outperform expert systems in most settings. These models are typically applied to features computed from customer consumption profiles such as average consumption, maximum consumption and change of consumption in addition to customer master data features such as type of customer and connection type. Sizes of databases used in the literature have a large range from less than 100 to more than one million. In this survey, we have also identified the six main open challenges in NTL detection: handling imbalanced classes in the training data and choosing appropriate evaluation metrics, describing features from the data, handling incorrect inspection results, correcting the bias in the inspection results, building models scalable to Big Data sets and making results obtained through different methods comparable. We believe that these need to be accurately addressed in future research in order to advance in NTL detection methods. This will allow to share sound, assessable, understandable, replicable and scalable results with the research community. In our current research we have started to create a database that we are planning to make available in the future. It will allow research groups to work on these challenges and and to assess their advancement in NTL detection.

References

[1] M. Abadi, A. Agarwal, P. Barham, et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems", 2015.

[2] M. S. Alam, E. Kabir, M. M. Rahman and M. A. K. Chowdhury, "Power Sector Reform in Bangladesh: Electricity Distribution System", Energy, vol. 29, no. 11, pp. 1773-1783, 2004.

[3] S. Amin, G. A. Schwartz and H. Tembine, "Incentives and security in electricity distribution networks", Decision and Game Theory for Security, Springer, pp. 264-280, 2012.

[4] E. W. S. dos Angelos, O. R. Saavedra, O. A. Carmona Cortes and A. Nunes de Souza, "Detection and identification of abnormalities in customer consumptions in power distribution systems", IEEE Transactions on Power Delivery, vol. 26, no. 4, pp. 2436-2442, 2011.

[5] E. Bair, "Semi-supervised clustering methods", WIREs Comp Stat, 5(5), pp. 349-361, 2013.

[6] K. Bottou, "Stochastic Learning", Advanced Lectures on Machine Learning, LNAI 3176, Springer, pp. 146-168, 2004.

[7] M. Banko and E. Brill, "Scaling to very very large corpora for natural language disambiguation", Proceedings of the 39th annual meeting on association for computational linguistics, pp. 26-33, 2001.

[8] B. Bhatia and M. Gulati, "Reforming the Power Sector. Controlling Electricity Theft and Improving Revenue", World Bank, 2004.

[9] C. M. Bishop, "Neural Networks for Pattern Recognition", Clarendon Press, 1996.

[10] C. M. Bishop, "Pattern Recognition and Machine Learning", Springer, 2007.

[11] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review Statistical science", Statist. Sci., vol. 17, issue 3, pp. 235-255, 2002.

[12] J. E. Cabral, J. O. P. Pinto, E. M. Gontijo and J. Reis Filho, "Rough Sets Based Fraud Detection in Electrical Energy Consumers", WSEAS Transactions on Mathematics, issue 2, vol. 3, pp. 413-416, 2004.

[13] J. E. Cabral, J. O. P. Pinto and A. M. A. C. Pinto, "Fraud detection system for high and low voltage electricity consumers based on data mining", Power & Energy Society General Meeting (PES'09), 2009.

[14] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines", ACM Transactions on Intelligent Systems and Technology, vol. 2, issue 3, pp. 27:1-27:27, 2011.

[15] A. Chauhan and S. Rajvanshi, "Non-Technical Losses in power system: A review", 2013 International Conference on Power, Energy and Control (ICPEC), pp. 558-561, 2013.

[16] C. Cortes and M. Mohri, "Domain adaptation and sample bias correction theory and algorithm for regression", Theoretical Computer Science, vol. 519, pp. 103-126, 2013.

[17] B. C. Costa, B. L. Alberto, A. M. Portela, W. Maduro and E. O. Eler, "Fraud detection in electric power distribution networks using an ANN-based knowledge-discovery process", International Journal of Artificial Intelligence & Applications, vol. 4, no. 6, 2013.

[18] E. Costa, F. Fabris, A. Rodrigues Loureiros, H. Ahonen and F. Miguel Varejao, "Optimization metaheuristics for minimizing variance in a real-world statistical application", SAC '13 Proceedings of the 28th Annual ACM Symposium on Applied Computing, pp. 206-207, 2013.

[19] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters", OSDI'04: Sixth Symposium on Operating System Design and Implementation, 2004.

[20] S. S. S. R. Depuru, L. Wang and V. Devabhaktuni, "Enhanced encoding technique for identifying abnormal energy usage pattern", North American Power Symposium (NAPS), 2012.

[21] S. S. S. R. Depuru, L. Wang and V. Devabhaktuni, "Support vector machine based data classification for detection of electricity theft", Power Systems Conference and Exposition (PSCE), 2011.

[22] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni and R. C. Green, "High Performance Computing for Detection of Electricity Theft", International Journal of Electrical Power & Energy Systems, vol. 47, issue 1, pp. 21-30, May 2013.

[23] P. Domingos, "A few useful things to know about machine learning", Communications of the ACM, vol. 55 issue 10, pp 78-87, October 2012.

[24] P. Glauner, A. Boechat, L. Dolberg, R. State, F. Bettinger, Y. Rangoni and D. Duarte. "Large-Scale Detection of Non-Technical Losses in Imbalanced Data Sets", *submitted to the* Seventh IEEE Conference on Innovative Smart Grid Technologies (ISGT 2016), arXiv:1602.08350, 2016.

[25] D. Gorgevik, D. Cakmakov and V. Radevski, "Handwritten digit recognition using statistical and rule-based decision fusion", 11th Mediterranean Electrotechnical Conference (MELECON), pp. 131-135, 2002.

[26] Tim Harford, "Big data: are we making a big mistake?", FT Magazine, March 28, 2014.

[27] G. Hinton, L. Deng, D. Yu, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, G. Dahl and B. Kingsbury, "Deep Neural Networks for Acoustic Modeling in Speech Recognition", IEEE Signal Processing Magazine, 29 (6), 82-97, 2012.

[28] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study", Intelligent Data Analysis, vol. 6, issue 5, pp. 429-449, 2002.

[29] R. Jiang, R. Lu, Y. Wang and J. Luo, "Energy-theft detection issues for advanced metering infrastructure in smart grid", Tsinghua Science and Technology, vol. 19, issue 2, pp. 105-120, 2014.

[30] M. Kazerooni, H. Zhu and T. J. Overbye, "Literature review on the applications of data mining in power systems", Power and Energy Conference at Illinois (PECI), 2014.

[31] S. L. Kendal and M. Creen, "An introduction to knowledge engineering", Springer, ISBN 978-1-84628-475-5, 2007.

[32] Y. Kou, C.-T. Lu, S. Sirwongwattana and Y.-P. Huang, "Survey of fraud detection techniques", IEEE International Conference on Networking, Sensing and Control, vol. 2, pp. 749-754, 2004.

[33] S. H. Kwak, "A mission planning expert system for an autonomous underwater vehicle", Proceedings of the 1990 Symposium on Autonomous Underwater Vehicle Technology, pp. 123-128, 1990.

[34] Y. LeCun, Y. Bengio and G. E. Hinton, "Deep Learning", Nature, vol. 521, pp. 436-444.

[35] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel, "Handwritten digit recognition with a back-propagation network", Advances in Neural Information Processing Systems 2 (NIPS '89), Denver, CO, 1990.

[36] L. J. P. van der Maaten and G. E. Hinton, "Visualizing High-Dimensional Data Using t-SNE", Journal of Machine Learning Research, 9 (Nov): 2579-2605, 2008.

[37] M. Di Martino, F. Decia, J. Molinelli and Alicia Fernandez, "Improving electric fraud detection using class imbalance strategies", 2012.

[38] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme", Biochimica et Biophysica Acta (BBA) - Protein Structure 405 (2): pp. 442-451, 1975.

[39] J. McCarthy, M. Minsky, N. Rochester and C. Shannon, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence", 1955.

[40] S. McLaughlin, D. Podkuiko and P. McDaniel, "Energy Theft in the Advanced Metering Infrastructure", Lecture Notes in Computer Science, vol. 6027, pp. 176-187, 2009.

[41] C. Muniz, K. Figueiredo, M. M. B. R. Vellasco, G. Chavez and M. A. C. Pacheco, "Irregularity detection on low tension electric installations by neural network ensembles", IEEE - INNS - ENNS International Joint Conference on Neural Networks, June 2009.

[42] C. Muniz, M. M. B. R. Vellasco, R. Tanscheit and K. A. Figueiredo, "Neuro-fuzzy System for Fraud Detection in Electricity Distribution", IFSA/EUSFLAT Conference, pp. 1096-1101, 2009.

[43] J. Nagi, A. M. Mohamad, K. S. Yap and S.K Tiong, "Non-Technical Loss analysis for detection of electricity theft using support vector machines", IEEE 2nd International Power and Energy Conference (PECon 2008), 2008.

[44] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed and A. M. Mohammad, "Detection of abnormalities and electricity theft using genetic Support Vector Machines", 2008 IEEE Region 10 Conference TENCON, 2008.

[45] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed and F. Nagi, "Improving SVM-Based Nontechnical Loss Detection in Power Utility Using the Fuzzy Inference System", IEEE Transactions on Power Delivery, vol. 26, no. 2, pp. 1284-1285, 2011.

[46] J. Nagi, K. S. Yap, S. K. Tiong, S. K. and A. M. Mohamad, "Non-technical Loss Detection for Metered Customers in Power Utility Using Support Vector Machines", IEEE Transactions on Power Delivery, vol. 25, no. 2, pp. 1162-1171, 2009.

[47] J. Nagi, K. S. Yap, F. Nagi, S. K. Tiong, S. P. Koh and S. K. Ahmed, "NTL detection of electricity theft and abnormalities for large power consumers In TNB Malaysia", 2010 IEEE Student Conference on Research and Development (SCOReD), pp. 202-206, 2010.

[48] W. R. Nelson, "REACTOR: An Expert System for Diagnosis and Treatment of Nuclear Reactor Accidents", Proceedings AAAI-82, 1982.

[49] A. Ng, "Feature selection, L1 vs. L2 regularization, and rotational invariance", Stanford, 2004.

[50] A. Ng, "Machine Learning", Coursera, 2014.

[51] A. Ng, A. Coates, M. Diel, V. Ganapathi, J. Schulte, B. Tse, E. Berger and E. Liang, "Inverted autonomous helicopter flight via reinforcement learning", International Symposium on Experimental Robotics, 2004.

[52] A. H. Nizar, Z. Y. Dong and Y. Wang, "Power Utility Nontechnical Loss Analysis With Extreme Learning Machine Method", IEEE Transactions on Power Systems, vol. 23, issue 3, pp. 946-955, 2008.

[53] A. H. Nizar, J. H. Zhao and Z. Y. Dong, "Customer information system data pre-processing with feature selection techniques for non-technical losses prediction in an electricity market", International Conference on Power System Technology (PowerCon 2006), 2006.

[54] C. C. Oba Ramos, A. Nunes Souza, J. Paulo Papa and A. Xavier Falcao, "A New Approach for Nontechnical Losses Detection Based on Optimum-Path Forest", IEEE Transactions on Power Systems, vol. 26, issue 1, pp. 181-189, 2011.

[55] C. C. Oba Ramos, A. N. de Souza, J. P. Papa and A. X. Falcao, "Fast Non-Technical Losses Identification Through Optimum-Path Forest", 15th International Conference on Intelligent System Applications to Power Systems (ISAP), November 2009.

[56] C. C. Oba Ramos, A. Nunes Souza, J. Paulo Papa and A. Xavier Falcao, "Learning to Identify Non-Technical Losses with Optimum-Path Forest", 17th International Conference on Systems, Signals and Image Processing (IWSSIP 2010), 2010.

[57] C. C. Oba Ramos, A. Nunes de Souza, D. Sinkiti Gastaldello and J. Paulo Papa, "Identification and feature selection of non-technical losses for industrial consumers using the software WEKA", International Conference on Industry Applications, 2012.

[58] C. C. B. de Oliveira, N. Kagan, A. Meffe, S. L. Caparroz and J. L. Cavaretti, "A New Method for the Computation of Technical Losses in Electrical Power Distribution Systems", Proceedings ClRED, 2001.

[59] R. Penrose, "A generalized inverse for matrices", Proceedings of the Cambridge Philosophical Society 51, pp. 406-413, 1955.

[60] L. A. M. Pereira, L. C. S. Afonso, J. P. Papa, Z. A. Vale, C. C. O. Ramos, D. S. Gastaldello and A. N. Souza, "Multilayer perceptron neural networks training through charged system search and its Application for non-technical losses detection", Conference On Innovative Smart Grid Technologies Latin America (ISGT LA), 2013.

[61] J. Peters, S. Vijayakumar and S. Schaal, "Reinforcement Learning for Humanoid Robotics", IEEE-RAS International Conference on Humanoid Robots, 2003.

[62] S. Russel and P. Norvig, "Artificial Intelligence: A Modern Approach", Prentice Hall, Third Edition, 2009.

[63] S. Sahoo, D. Nikovski, T. Muso and K. Tsuru, "Electricity theft detection using smart meter data", IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2015.

[64] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel and D. Hassabis, "Mastering the game of Go with deep neural networks and tree search", Nature, vol. 529, pp. 484-489, 2016.

[65] T. B. Smith, "Electricity theft: a comparative analysis", Energy Policy, vol. 32, issue 18, pp. 2067-2076, December 2004.

[66] J. V. Spiric, S. S. Stankovic, M. B. Docic and T. D. Popovic, "Using the rough set theory to detect fraud committed by electricity customers", International Journal of Electrical Power & Energy Systems, vol. 62, pp. 727-734, November 2014.

[67] R. S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction", MIT Press, 1998.

[68] Y. Tang, Y.-Q. Zhang, N. V. Chawla and S. Krasser, "SVMs Modeling for Highly Imbalanced Classification", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 39, issue 1, pp. 281-288, Feb. 2009.

[69] V. N. Vapnik, "An overview of statistical learning theory", IEEE Transactions on Neural Networks, vol. 10, issue 5, pp. 988-999, Sep. 1999.

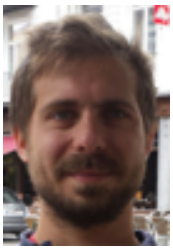
[70] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker and I. Stoica, "Spark: cluster computing with working sets", HotCloud'10 Proceedings of the 2nd USENIX conference on Hot topics in cloud computing, 2010.

[71] Z. Zeng, M. Pantic, G. I. Roisman and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, issue 1, pp. 39-58, 2009.

Patrick Glauner (M'12) received the B.Sc. degree in computer science from Karlsruhe University of Applied Sciences in 2012 and M.Sc. degree in machine learning from Imperial College London in 2015. He was a Fellow at CERN, the European Organization for Nuclear Research, worked at SAP and is an alumnus of the German National Academic Foundation (Studienstiftung des deutschen Volkes). He is currently a Ph.D. student in machine learning in the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, under the supervision of Dr. Radu State. His interests include artificial intelligence, machine learning, deep learning, anomaly detection and big data.

Andre Boechat is a Computer Engineer and, for the last four years, had worked with machine learning, image processing, OCR systems, distributed systems and web spiders. Currently, as a PhD candidate, he is working with detection of abnormalities in customer consumptions in power distribution systems.

Jorge Meira was born in Brazil in 1983. He received the Ph.D. degree in computer Science from the University of Luxembourg, Luxembourg, in 2014. He has held an Assistant Professor position at Federal University of Paran, Brazil, in 2015. His main areas of research interest are Software Testing, Databases Systems and Big Data. He is currently a Research Associate at University of Luxembourg.



Lautaro Dolberg was born in Argentina, were he did his master in Computer Science and worked as a software engineer for 7 years. After doing an internship at INRIA, France he moved to Luxembourg to pursue a PhD in Informatics at the University of Luxembourg, his thesis was entitled Network Management with Data Analytics. Currently, he works as Data Scientist consultant while doing a Postdoc at the University of Luxembourg and Choice Holdings.

Radu State is a senior scientist at SnT and heads the research group SEDAN (Service and Data Management in Distributed Systems). He holds a Master of Science degree from the Johns Hopkins University, USA and a PhD degree otbtained during his research activity with INRIA, France. He was a Senior Researcher at INRIA, France and Professor of Computer Science at Telecom Nancy, France.

Franck Bettinger was born in France in 1976. He holds an engineering degree from the Institut d'Informatique d'Entreprise (ENSIIE) in Evry, France, and a M.Sc. in Advanced Computer Science from the University of Manchester, UK. He received a Ph.D. degree for his work on time varying active appearance model, in 2004, from the University of Manchester, UK. He worked as a software engineer for 10 years in Luxembourg and is now working for Choice Technologies Holding on electricity fraud detection problems. His interests cover machine vision, machine learning in general, and robotics.

Yves Rangoni received his M.Sc. in Computer Vision and was awarded with a Ph.D. degree in Computer Science from Nancy 2 University, France, for his work on Document Image Analysis and Recognition and Neural Networks at the Loria Research Center in 2007. Then, he worked at the DFKI Kaiserslautern, Germany, on Optical Character Recognition topics, at the INRIA, France, on Logical Layout Analysis, at the Tudor Research Center, Luxembourg, on Handwriting Recognition on Tangible User Interfaces. He is now a research engineer at Choice Technologies Holding company Luxembourg, working on fraud detection problems.

Diogo Duarte was born in Brazil in 1980. He received the M.Sc. degree in Electrical Engineering from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio), and B.Sc. degree in Electrical Engineering from the State University of Rio de Janeiro (UERJ), both in Brazil. He is a senior software engineer in Choice Technologies, working in different R&D projects in Brazilian energy market. His research interests include statistical pattern recognition, neural networks, computer vision, software development, databases and energy losses.