Improving Knowledge-Based Systems with statistical techniques, text mining, and neural networks for non-technical loss detection

Juan I. Guerrero *, Carlos León, Iñigo Monedero, Félix Biscarri, Jesús Biscarri

Department of Electronic Technology, Universidad de Sevilla, ETSII, Av. Reina Mercedes, 41011 Seville, Spain

ABSTRACT

Currently, power distribution companies have several problems that are related to energy losses. For example, the energy used might not be billed due to illegal manipulation or a breakdown in the customer's measurement equipment. These types of losses are called *non-technical losses* (NTLs), and these losses are usually greater than the losses that are due to the distribution infrastructure (technical losses). Traditionally, a large number of studies have used data mining to detect NTLs, but to the best of our knowledge, there are no studies that involve the use of a Knowledge-Based System (KBS) that is created based on the knowledge and expertise of the inspectors. In the present study, a KBS was built that is based on the knowledge and expertise of the inspectors and that uses text mining, neural networks, and statistical techniques for the detection of NTLs. Text mining, neural networks, and statistical techniques were used to extract information from samples, and this information was translated into rules, which were joined to the rules that were generated by the knowledge of the inspectors. This system was tested with real samples that were extracted from Endesa databases. Endesa is one of the most important distribution companies in Spain, and it plays an important role in international markets in both Europe and South America, having more than 73 million customers.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Losses are a serious issue facing utility distribution companies. In power distribution companies, there are two types of losses: technical losses and non-technical losses (NTLs).The NTLs are caused by illegal manipulations or faults in consumer facilities, and the amount of consumption cannot be registered by the system, whereas technical losses are caused by physical effects (e.g., the Joule effect) that are due to the power distribution. The technical losses can be forecasted with a low error rate and might be reduced by means of using better facilities. However, the reduction of NTLs is a more difficult problem because it is necessary to detect the frauds and anomalies in the consumer facilities. Thus, the use of data mining and computational intelligence techniques for NTL detection has grown exponentially over the past decade.

In addition, some of the new technologies are related to Smart Grids [1]. These technologies provide new meter infrastructures [2] and new software infrastructures [3], which increase the information about consumers and the service that distribution companies can offer them. One of the most popular services is the application of enterprise information systems that are based on cloud or distributed management [4].

The European Commission [5] defined a Smart Grid as follows: "A Smart Grid is an electricity network that can cost efficiently integrate the behaviour and actions of all users connected to it – generators, consumers and those that do both – to ensure economically efficient, sustainable power system with low losses and high levels of quality and security of supply and safety". This definition is also based on [6,7], but it is possible to find other references that provide some additional information about the Smart Grid paradigm ([8–11], etc.). The benefits that are associated with Smart Grids include the following:

– Better integration of customer-owner power generation systems.
– Better integration of renewable energy systems.
– More efficient transmission of electricity.
– Quicker restoration of electricity after power faults and disturbances.
– Reduced operations and management costs for utilities.
– Reduced power costs for consumers.

* Corresponding author.
E-mail addresses: juaguealo@us.es (J.I. Guerrero), cleon@us.es (C. León), imonedero@us.es (I. Monedero), fbiscarri@us.es (F. Biscarri), jbiscarri@us.es (J. Biscarri).

- Reduced peak demand, in turn, will decrease the electricity rates.
- Improved security.
- Better integration of electrical vehicles.

There are several references about the advantages of Smart Grids, and there are many initiatives that are related to Smart Grids all over the world; it is possible to find reports that describe different initiatives around the world (e.g., [12–14]).

The Advanced Metering Infrastructure (AMI) provides new technologies that improve the capabilities of the power registers, using smart meters. The AMI is an important part of the Smart Grid and provides a high-level control application over the power grid. There are several studies about the utilisation of AMIs in Smart Grids (e.g., [15–17]) to improve the power quality (e.g., [18,19]) and demand management (e.g., [20,21]).

These new technologies provide new resources for researching and development projects, providing new software infrastructure that is related to applications such as advanced SCADA (Supervisory Control and Data Acquisition) [22], energy management systems [23], and demand forecasting [24], among others. All of these software and hardware approaches are coordinated in a distributed power grid, providing a more efficient power distribution grid.

Currently, due to the installation of smart technologies, both modern and old facilities coexist in the same grid. The company inspector dedicated to the detection and correction of NTLs has the necessary knowledge to analyse the consumer information. In addition, this new arrangement causes a change in the business model of the distribution companies. Currently, utility distribution companies are assuming outsourcing models. The outsourcing is the contracting out of a business process to a third party and sometimes involves transferring employees and assets from one company to another ([25] provides a study about this business model). These are the main reasons why the power distribution companies and utilities in general are interested in research on automatic methods for NTL detection and for preserving the knowledge of the inspectors, who have a large amount of experience working at the company.

This article shows a Knowledge-Based System (KBS) that is created using the knowledge provided by the inspectors of the distribution company. This KBS increases their efficiency using statistical techniques, text mining and neural networks. Endesa, which is one of the most important companies of Spain, supplies energy to approximately 73 million consumers in South America and Europe, and it provides inspectors who possess knowledge and test the developed system.

Traditionally, the procedure of NTL detection is performed on a very large number of consumers (several thousand) using inspections. This procedure involves a 'manual analysis' of the information that is stored in the databases of the company to determine whether an NTL could exist. The 'manual analysis' is performed by means of several query tools that are provided by the company, and the inspectors must review and study the results of each tool for each consumer. Although the inspectors have a large amount of experience, this process usually takes a long time. This approach is the main technique that the inspectors typically perform to increase their efficiency. Simplifying this process by means of a KBS increases the efficiency of the inspections and also facilitates the training of new staff.

A description of the proposed KBS is given in this article. First, the primary studies that are related to this system are reviewed. Second, the knowledge-based development is described. Third, the architecture of the system and its implementation are shown. Finally, the experimental results in real cases and the conclusions are described.

2. Bibliographical review

Most of the consulted references describe techniques that are based on the upgrading of the distribution infrastructure and/or the use of supervised learning techniques. The supervised techniques use cases that have been previously detected to obtain the corresponding patterns.

Aguero [26] proposed the improvement of the efficiency of the distribution systems in reducing the technical and non-technical losses. A utility information system that includes computational models of feeders and advanced modelling software systems is used. However, this system is based on the implementation of the approach of using smart grids. In the same way, Paruchuri and Dubey [27] proposed the use of smart metering and advanced communication protocols to detect NTLs. Iglesias [28] proposed an analysis of energy losses for sectors using a load balance, by means of the consumer's information, the distribution transformers and several measurement points. Alves et al. [29] suggested an upgrading of the measurement equipment by means of electronic devices such as alarm systems, connection systems, remote reconnection, and protection of the drivers.

Nizar et al. [30] described a series of detection rules that are based on feature selection and clustering techniques, using the consumption of the consumers. Depuru et al. [31] proposed the use of consumption patterns that are generated starting from the historical consumption of consumers, using a Support Vector Machine (SVM), and the authors gathered data from smart meters; the consumers are classified according to the pattern that they present. Ramos et al. [32] proposed the use of an Optimum Path Forest (OPF) clustering technique based on supervised learning; thus, a dataset that was obtained and included fraudulent activity from a power distribution company was used. These references are based on the use of learning techniques using available information about the consumption of consumers.

The consulted references use 'data-driven methods' based on the automatic detection of the NTL pattern. These references do not directly include the knowledge of the inspectors in the NTL detection process, except in the case of the confirmation of the results of the inspections for validation purposes.

Traditionally, Knowledge-Based Systems (KBS) are used in power systems control [33], and they are classified into three important components: generation, transmission, and distribution. Mostly, this type of KBS is combined with other types of techniques. For generation, Yang and Chang [34] proposed a diagnostic expert system for a nuclear power plant based on the hybrid knowledge approach called HYPOSS (hybrid knowledge-based plant operation support system). Zhang [35] described a frequency and knowledge tree/casualty diagram-based expert system approach for the on-line/off-line fault diagnoses of engineering systems. Ross et al. [36] proposed a KBS for the scheduling of an energy storage system (ESS) that was installed in a wind-diesel isolated power system; by decreasing the energy that is wasted through the dump load with the use of the ESS and KBES (Knowledge-based Expert System) controller, the required diesel generation, operation costs and emissions were reduced. For transmission, Yuehui and Yihan [37] developed a knowledge- and network-based cognition model expert system of automatic switching sequences for substation and power plants, and this system was successfully used by a number of substation and power plants in China. Picard et al. [38] proposed a Knowledge-Based System for the structural design of high-voltage lines, and the basic structure of the system consisted of a database of tower configurations and associated parameters, providing the cost per kilometer of a high-voltage transmission line. For distribution, Kandil et al. [39] used a KBES to support the choice of the most suitable load forecasting model for medium- and

long-term power system planning; Reddy and Sydulu [40] proposed an algorithm for network reconfiguration based on a heuristic-expert approach; and Abdelaziz et al. [41] developed a fuzzy expert system for loss reduction and voltage control in radial distribution systems, minimising the total system losses. There are no references about NTL detection using KBS in power distribution databases. However, there are different references about fraud detection in other research areas; for example, Šubelj et al. [42] showed an expert system for detecting automobile insurance fraud; Hilas [43] presented an expert system for fraud detection in private telecommunication networks; Leonard [44] proposed an expert system to detect credit card fraud; and Leonard [45] developed a rule-based expert system model for fraud alert in consumer credit.

Currently, most of the references are about the application of data-driven techniques to NTL detection. There are not many articles about knowledge-driven solutions because it is unusual and expensive that a distribution company provides access to their inspectors. The data-driven techniques are the most used because data and some knowledge about it are needed, and data-driven techniques are relatively recent.

Although the proposed system has two subsystems (statistical analysis of the consumption, text mining and neural networks) with data-driven models, the main core and the knowledge base are built using knowledge from inspectors. Even those subsystems with data-driven models were inspired with the concepts and processes described by inspectors in the knowledge acquisition sessions.

The KBS proposed in this article was developed into Project MIDAS. There are previous studies about project MIDAS, which are described in [46,47]. First, in [46], the different methods applied in the MIDAS project for NTL detection were described. The remainder of this paper is about data mining techniques that use information about consumption and some other characteristics from contracts, and these techniques are described in [48,49]. These references provided several data-driven models with a different level of inspectors' knowledge. The quantity of information that is involved in the analysis of the consumers is another difference between the data mining techniques and the proposed KBS. There is a rule-based expert system described in [46]; this RBES is described in [47] in detail. In this reference, an expert system was combined with statistical techniques, text mining, and neural networks to analyse information about the consumers. This RBES is based in a series of rules that were applied in a procedural way with some exceptions. However, in the proposed KBS, the system requires a rule engine that is provided by COOL because the inspector's knowledge on which the KBS is based comes from trainers of inspectors and they provided high quality knowledge that is more complex to implement in a procedural system. This complexity arises because the process increased the number of parameters and characteristics about the consumer that are involved in the process. In addition, this new knowledge provided rules that allowed us to detect new types of NTLs, which the RBES (proposed in [47]) could not detect. A description of the text mining and neural networks was proposed in [50]. A new version of this system is described in this article, improving the number of semantic categories and the number of concepts (covering languages: Catalan, Aranese, and Balearic). Mainly, the new version includes several additional rules that were acquired from the inspectors of Endesa; the statistical techniques, with a normalisation process for consumption, text mining, and neural networks, include a large quantity of new concepts from inspectors' commentaries and comprise a new module that translates all of the information generated by statistical techniques, text mining, and neural networks to instances of COOL objects.

3. Knowledge acquisition procedure

The inspectors who work at power distribution companies possess immense knowledge about the detection of NTLs. The inspectors usually perform a manual analysis of a series of data on consumers.

A typical NTL detection procedure is commonly composed of the following steps:

1. Manual analysis of the data on the contracts of consumers.
2. Manual analysis of the consumption of the consumers according to their characteristics.
3. Manual analysis of comments contributed by other inspectors in other inspections, or if it is about a customer without smart metering, they also review the comments of company staff (if there are any), who pick up the information from the measurement equipment.
4. Manual analysis of previous NTLs detected.

The analysis is conducted with larger or smaller depths depending on the available information about each consumer. This process is performed for each consumer and takes a long time. At the same time, the execution of this analysis improves the final efficiency of the inspections because some inspections can be avoided.

In this article, a KBS is proposed. Knowledge from the inspectors was used to develop the system that performs the process of the manual analysis. Knowledge engineering was applied to gather information. This approach included different techniques that enable the acquisition of knowledge from inspectors.

The inspectors' experience can usually influence the success of acquisition processes. Sometimes, the inspectors have a high level of experience; in this case, from the point of view of the inspectors, the procedure can be trivial and thus, it is difficult to convey the result because the inspectors could omit important details that they consider trivial.

The procedures of knowledge acquisition used in this case are as follows:

– The personal interview. This interview consists of a series of questions that are identified with the problem domain. These questions can be about general or specific topics. The inspectors should be previously prepared with this technique, marking the objectives in each session. The knowledge engineer should have the capacity to drive the interview correctly, to avoid ambiguities and loss of the objective.
– The structured interview. This technique consists of a set of interviews that are driven by objectives. This type of question allows the participants to synthesise their knowledge with respect to a specific task. This technique has been demonstrated to be more efficient than the other methods that are used in the knowledge extraction process. However, this technique requires a long time to prepare the session because the knowledge engineer should make questions that allow the participants to discover new concepts in the problem domain, to cover the relationships and to avoid ambiguities.
– The technique of the observation of experts. This technique consists of examining the work of experts at their work site. The main features that were accumulated through this technique were related to the decisions and processes that the experts conducted.
– The technique of protocol analysis. This technique is very useful in all environments; with this technique, the knowledge engineer designs a possible scenario, and the expert exposes the solution process (including queries, decisions). Thus, this process was structured, and it allowed obtaining a performance protocol for certain analysis scenarios.

After these interviews, a summary of the acquired knowledge was written and sent to the inspectors for their approval. Last, the translation and implementation of the knowledge were developed.

In addition, each conclusion that was extracted from each interview was verified with the other inspectors who were involved in the process. In this way, several inspectors reviewed the knowledge from the previously interviewed inspectors.

The inspectors who are contracted to Endesa usually train the inspectors of the outsourced companies. The criteria for the selection of the experts or inspectors were the following:

- The inspectors must be contracted by the company.
- They must have a high level of experience with manual analysis and inspections.
- They must have a high ratio of NTL detection.
- Endesa works in five different zones in Spain. Three of them are small, one is of medium size, and the last is large. For each zone, a different number of experts is selected: one inspector is selected from each of the small zones, two inspectors are selected from each of the medium-size zones, and three inspectors are selected for the large zone.

Several support indexes were assigned to each inspector of Endesa. The support index is calculated as follows:

- Career. The inspector's career in years; if the inspector is not contracted within the company, this value will be 0.
- Analysis_experience. The number of customers manually analysed per year.
- Inspection_experience. The number of customers inspected per year.
- Success_ratio. The number of customers analysed and inspected who finally had an NTL divided by the total number of customers analysed and inspected.
- Zones. The number of zones where they have worked.

In this way, eight inspectors were selected for the knowledge acquisition process, according to these support indexes.

The knowledge was acquired in the following manner:

- Each inspector was given a personal interview. The result of this session was validated by the corresponding inspectors and by the other inspectors. In total, eight personal interview sessions of eight inspectors were performed.
- Each inspector was also observed in an observation session. In total, eight observation sessions of eight inspectors were performed.
- One inspector, who had the highest level of experience, was interviewed in a structured interview. In total, one structured interview was performed.
- One inspector, who had the highest level of experience, was interviewed in a session on protocol analysis. In total, one protocol analysis session was performed.

All of these processes provided the creation of a KBS that uses statistical techniques and text mining to supplement the knowledge of the inspectors.

3.1. Personal and structured interviews

Personal and structured interviews are the main techniques that were applied to the acquisition of knowledge from inspectors. These techniques required a database of questions. This database gathered approximately 600 questions about different topics of the consumer analysis process. These questions are structured into groups and subgroups. The main groups of questions are the following:

- Group A. Contact information.
- Group B. Information about the work area (geographical locations) in which the inspector usually works, namely, the quantity of detected NTL against the quantity of manually analysed consumers.
- Group C. Information about NTL that is usually found in the work area of the inspector.
- Group D. Information about the relationship between the information stored in the database and the real information that exists in the consumer facilities.
- Group E. Information about the procedures that the inspectors usually perform.
- Group F. Specific questions.

In the case of the structured interviews, these questions are related to previously identified supplies and the procedures that identified special supplies.

A questionnaire is written for each inspector. In the case of personal interviews, the new ideas obtained in each interview are included in the next questionnaire for peer review. The inspectors specified only small variations between the behaviours of different zones. Thus, the number of questions that were added from previous interviews decreased to one or two after the second interview. In the case of structured interviews, the questions are organised depending on the answers and the case selected for the session.

3.2. Non-technical losses

Non-technical losses (NTLs) are caused by illegal manipulations or breakdowns in the consumer's measurement equipment, and the consumption cannot be registered by the system. Thus, although the energy was consumed by the consumer, it was not registered or billed. The power distribution companies usually identify three different types of NTLs:

- Anomalies without demonstrable energy losses. The consumer's measurement equipment might be influenced by the anomaly. There will not be any new invoice, because the inspector cannot demonstrate whether it will be caused by a consumer or the company staff.
- Anomalies with energy losses. This case is the same case as the anomalies without demonstrable energy losses, but, in this case, the inspector can estimate the quantity of energy that was not registered by the power register. It is possible to open a proceeding to return to normal the consumer's measurement equipment.
- Frauds. Fraud is the illegal manipulation of measurement equipment for the purpose of avoiding the appropriate amount of consumption registration in the power register. In this case, the inspectors open a proceeding against the consumer. Then, the consumer can negotiate with the inspectors to determine the disciplinary action by means of a fine.

The energy that is consumed, billed and not paid for is an administrative problem. This is not considered to be an NTL in the distribution companies, because the energy consumption is registered by the system. In Spain, the power sector was split up into commercialisation and distribution because of a 2009 law about the liberalisation of the power supply services. Thus, the commercialisation company is responsible for this administrative

problem. However, these cases are automatically controlled in the commercialisation companies. If it is needed, a work order is automatically sent to the distribution company to perform a power cut-off in the consumer supply.

In addition, the application of these new technologies provides some advantages that allow us to reduce the non-technical losses. However, in the NTL detection area, new types of NTLs will appear that are related to communication anomalies, faults, or frauds. In fact, the main difference between smart meters and traditional meters is the measurement capability. Usually the problems that are related to telecommunication in the smart meter are treated as an anomaly without demonstrable energy losses.

The consumers that are involved with low voltage usually have traditional meters. They usually have monthly or bimonthly meter readings, and sometimes these one are estimated because the power register is inside the building and it is not possible to take a reading. The analysis (both manual analysis and KBS analysis) is influenced by these problems, and in some cases, it is not possible to make a decision about the existence of an NTL at the consumer site.

4. Knowledge-Based System (KBS)

The KBS has been implemented using IBM Modeler and CLIPS (C Language-Integrated Production System) or COOL (CLIPS Object Oriented Language). The IBM SPSS Modeler is a software used for data mining processes; it provides several tools for treating different information types and several methods that are used for inference knowledge. Initially, this software system was named SPSS Clementine, but the name was changed to IBM SPSS Modeler when IBM bought SPSS. CLIPS is a software system that is used to create and run rule-based expert systems.
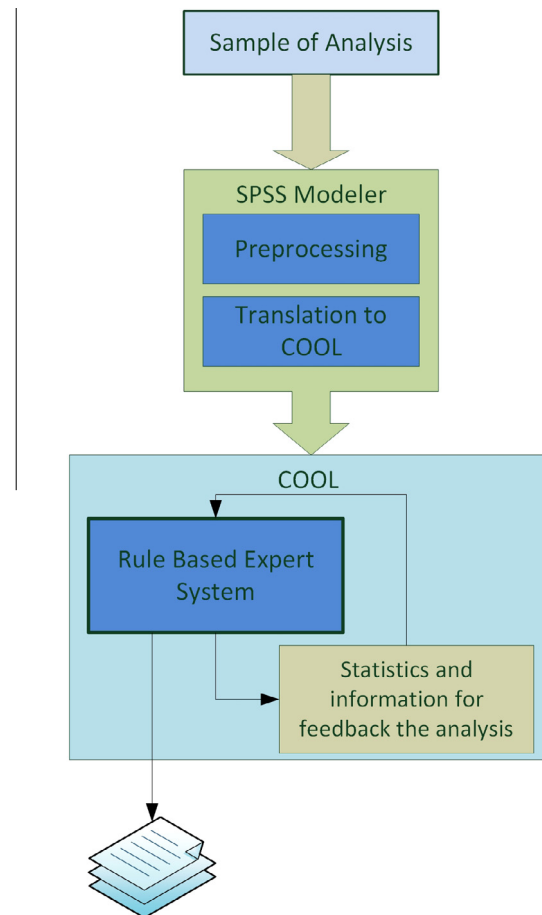
The analysis procedure was identified according to the process of acquisition, and the NTL detection rules were identified. The manual analysis requires very broad knowledge about company tools and power consumption. However, the expert systems usually have a restricted application domain. Nevertheless, the use of additional techniques allows an enlargement of this domain that the system can attain, and in turn, a new Knowledge-Based System with more capabilities is achieved.

The functional structure of the proposed KBS is shown in Fig. 1. The IBM SPSS Modeler and COOL systems are simultaneously performed; in this way, the system, when in a normal workflow, is preprocessing and analysed simultaneously.

4.1. Preprocessing

The databases of the power distribution companies have a large quantity of information that is not necessary for the analysis. There are between 500 Kb and several Mb per consumer. Thus, for 100,000 consumers, the information can be approximately 20 Gb to 50 Gb. This information is grouped in tables that have 3–90 columns. Each table stores information about different topics that are related to the consumers, including contract, consumption, and installation characteristics. This arrangement is one of the main reasons behind having a preprocessing step. The preprocessing step mainly involves the following operations:

  – Checking the coherence of the information.
  – Calculation of the different indicators that are used in the analysis.
  – Normalisation of some of the variables.
  – Discretisation of some of the variables.
  – Translation of the information to instances of the COOL classes.

Fig. 1. Chart of the functional structure of KBS.

The Sample of Analysis is usually formed from different sources. These sources are, for example, the databases of the different management systems, the text files, and excel (.xls) files.

The translation of information is made by means of instances of the COOL classes. These classes represent different aspects of the power distribution. The class diagram, which is shown in Fig. 2, shows the relations among the different COOL classes. If one class is connected to another, then there are one or more attributes in common; this relation is called inheritance. For example, the POWER class (Fig. 2) inherited the attributes and methods of the UTILITY class. This capability is one of the most important properties of object oriented programming.

In the class diagram, several types of classes are shown:

  – Classes that represent different scopes of utility (Common, Gas, Power, Consumption, and Distribution classes in Fig. 2). All of these are abstract classes.
  – Classes that represent the equipment of the grid (Connection, TC, GCP, and Country-Property classes in Fig. 2). All of these are inherited from the Distribution class.
  – Classes that represent the entities for consumption management (Contract, Supply, His-Measures, Client, Account, Contacts, His-Measure-Equip, Claims, Measure-Fault, Measure-Fault-Comment, and Claims-Comment classes in Fig. 2).
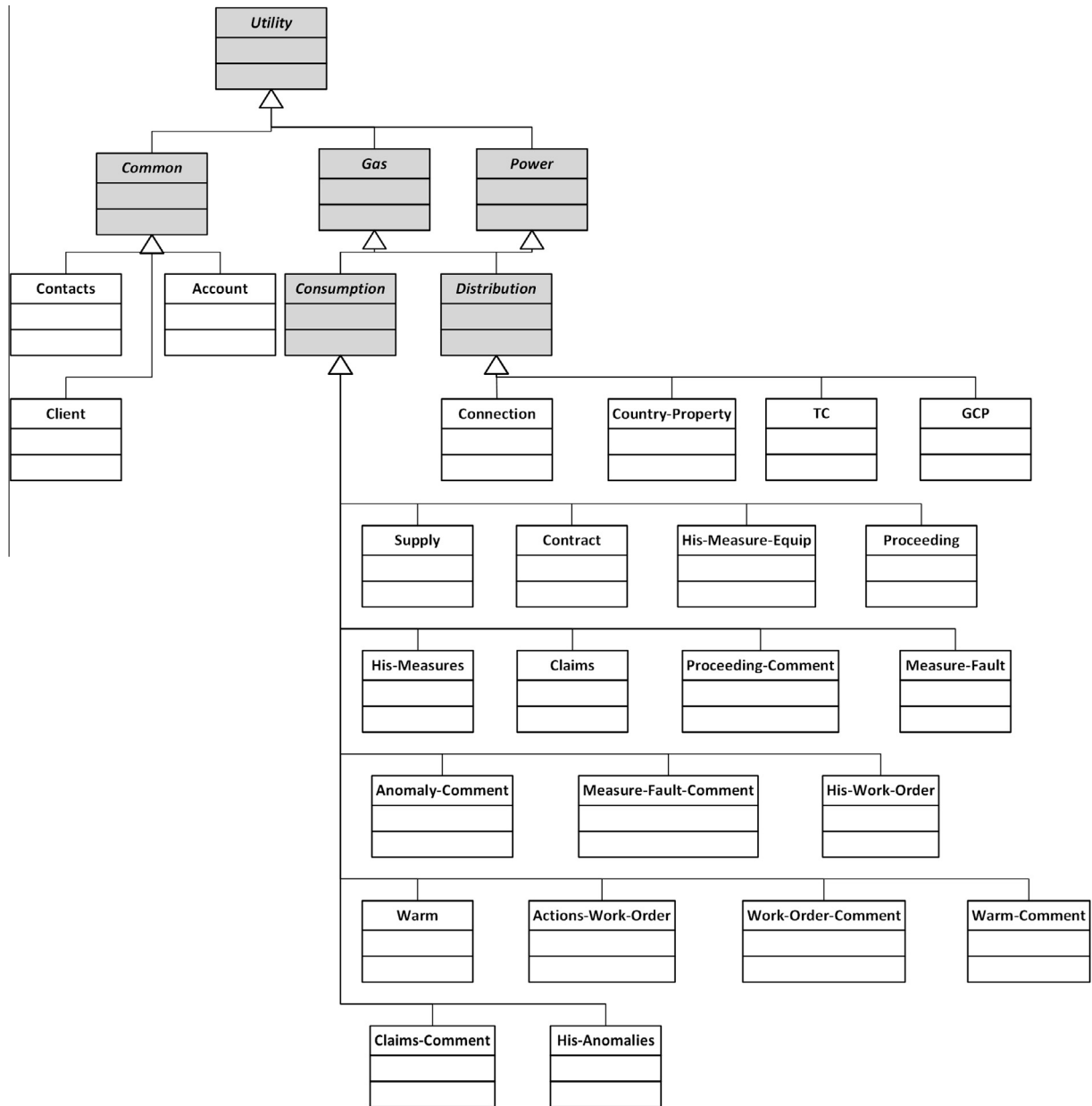
Fig. 2. Class diagram used to represent the entities in COOL.

– Classes that represent operations over the grid and consumer installations (Proceeding, Warm, His-Work-Order, His-Anomalies, Proceeding-Comment, Anomaly-Comment, Actions-Work-Order, Work-Order-Comment, and Warm-Comment classes in Fig. 2).

The translation process generates source code for the instances that are necessary to represent all of the information about the consumers. This information is used by rules when the analysis is performed in the COOL environment.

4.2. Statistical analysis of the consumption

In the manual analysis process, the inspectors conduct evaluations to determine whether the consumption of the client is correct. The interviews showed that the inspectors kept in mind the contracted power, the geographical location, the economic activity and the season. This procedure cannot be easily described by the

inspectors. The statistical analysis module attempts to imitate this process. The statistical analysis establishes the 'normal' pattern of client consumption according to the contracted power, the geographical location, the economic activity and the measurement date. These patterns were built with consumers who did not have NTLs. The clients who did not have NTLs are considered as 'normal' consumption patterns. The 'normal' pattern of consumption is defined with basic statistical indicators, including the maximum, minimum, average, and statistical deviation, according to the parameters that were previously mentioned. If a consumer fits into any of these patterns, this consumer will most likely not have an NTL.

The construction of these patterns are conducted with the following steps, which are shown in the flow diagram of Fig. 3. As shown in the diagram, the main indicators are the average, maximum, minimum, and typical deviation of the consumption, but they are calculated in several groups that correspond to the economic activity, geographical location, billing period, time discrimination

band, and time (monthly, annual, seasonal, and absolute). The statistical indicators are applied over normalised consumptions.

These patterns are modelled and generated with all of the information that is available about the consumers. The generated patterns are used in the analysis, and in turn, the patterns are updated by the KBS as a final step of each analysis. The larger the number of consumers that are analysed in the final step of the analysis, the more meaningful the statistical feedback will be. This process is initially conducted by IBM SPSS Modeler, but the updating process is conducted with rules that are implemented in COOL.

All of this information is translated to COOL instance objects, which will be used in the analysis as a part of the rules. The following COOL example shows the statistics of the monthly consumption of all of the customers in the sixth month or June (the consumption slot contains several numbers: the maximum, minimum, average, and standard deviation):

```
([gen19317] of ST-MONTHLY
    (consumption 109376.0 1885.793 21530.0 4290.398)
    (dt-study [EC-2-5-2011])
    (measurement-period "MONTHLY")
    (month 6)
)
```

[gen19317] is a random identifier for this instance of ST-MONTHLY, and the dt-study slot is the date of the last update that provides the results shown in the consumption slot.

4.3. Text mining and neural networks

In the manual analysis, the inspectors review the previous comments that are reported by other company inspectors or staff. This approach is the best format with which to determine the information about the consumer facilities and is one of the most important steps in the manual analysis. Text mining allows us to analyse the unstructured information (or text), providing rules for its analysis.

Natural Language Processing (NLP) techniques are used to analyse unstructured information, which can obtain a set of concepts [51]. The concepts represent words or groups of words that represent an idea or action. The concepts were extracted from the inspectors' commentaries. This process produced several million concepts and provided additional information about the frequency and percentage of appearance. This process is implemented and performed in SPSS Modeler. The NLP process is composed of four engines:

- String matching engine. This engine is based on fuzzy logic and uses synonym dictionaries. A fuzzy ratio is added to each word to identify similar words and errors. The error correction can be applied according to the lengths of the words.
- Syntactic engine. This engine assigns a function to each word, according to its position and the previous and following words.
- Concept engine. This engine generates several concepts; the words with the same syntactic function and meaning in the same sentence are grouped into the same concept.
- Semantic engine. This engine assigns a semantic function to each concept. The different semantic categories are defined according to the inspectors' knowledge.

The different languages were extracted by means of synonym dictionaries; these dictionaries allowed the inclusion of different languages in the extraction process. The extraction process involves the application of fuzzy techniques for string matching and the correction of errors.

After the extraction step, each concept was classified into a semantic category. The semantic category describes the purpose
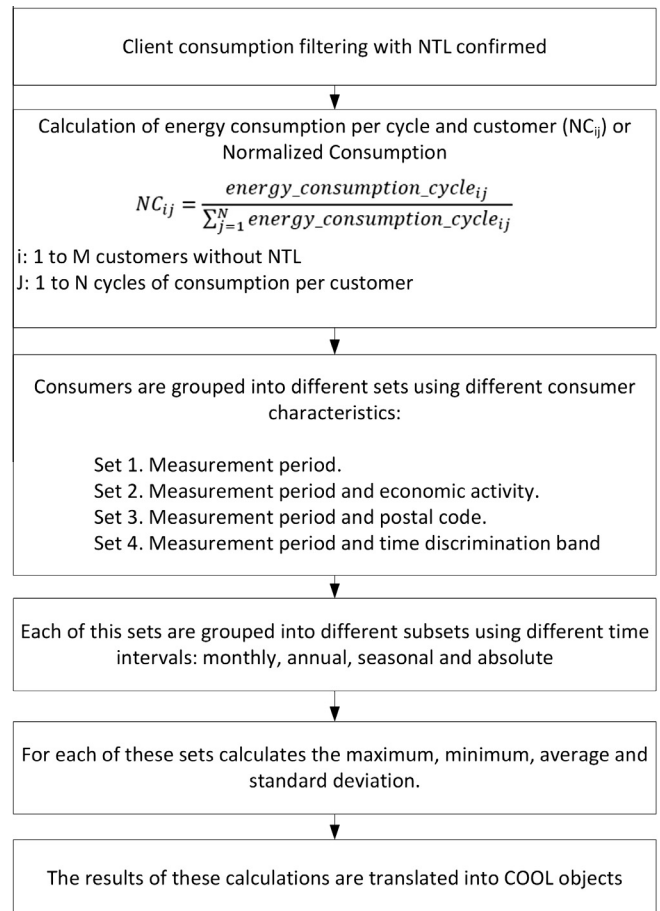

| Client consumption filtering with NTL confirmed |

Calculation of energy consumption per cycle and customer (NC$_{ij}$) or Normalized Consumption

$$NC_{ij} = \frac{energy\_consumption\_cycle_{ij}}{\sum_{j=1}^{N} energy\_consumption\_cycle_{ij}}$$

i: 1 to M customers without NTL
J: 1 to N cycles of consumption per customer

Consumers are grouped into different sets using different consumer characteristics:

Set 1. Measurement period.
Set 2. Measurement period and economic activity.
Set 3. Measurement period and postal code.
Set 4. Measurement period and time discrimination band

Each of this sets are grouped into different subsets using different time intervals: monthly, annual, seasonal and absolute

For each of these sets calculates the maximum, minimum, average and standard deviation.

The results of these calculations are translated into COOL objects

Fig. 3. Flow diagram of the statistical analysis of the consumption.

or objective of the concept, which is related to the NTL detections. There are several semantic categories. They are as follows:

- CLOSED. Concepts that identify consumers whose supplies are closed, uninhabited, on a holiday, demolished, and so on.
- CORRECT. Concepts that identify consumers whose consumer installations are correct or without NTLs, based on previous inspections.
- INCORRECT. Concepts that identify consumers whose consumer installation could have an NTL. These concepts are determined by inspectors and by the process described next.
- CB. Concepts that determine the existence of a capacitor bank.
- LOCATION. Concepts that identify consumer contact information.
- LOW CONSUMPTION. Concepts that identify consumers who traditionally have a low or very low consumption. Some concepts included consumers with economic activity who only have consumption in pseudo-random time periods, according to the external variables.

Overall, 99 additional semantic categories identify the information about the real economic activity of the consumers. These semantic categories are used to identify the possible mistakes in the original consumer information.

In addition, the information about the statistical analysis of the consumption was added to each concept to create a concepts dictionary that has associated characteristics. This dictionary is a multi-word relevant expression dictionary.

In this dictionary, 1% of the most frequent concepts (roughly 200,000) were manually classified into these categories, according

to the information obtained in the knowledge acquisition sessions. In addition, 23% of the remaining concepts were dates, numbers, and names, which can be used as additional information according to the context of the inspectors' commentaries.

The set of the most frequent concepts was used as a training set in a neural network. The neural network was trained by means of the multiple method. This method generated multiple networks that were simultaneously trained. Thus, this method initialised each network, and all of the networks were simultaneously trained. When a stopping criterion is met for all of the networks, the network with the highest accuracy is selected to be the final model. All of the neural networks had the same size at the beginning, with the same number of neurons in the input layer and first hidden layer. When the network is trained, the number of neurons in the hidden layers increases according to the value of the Softmax function. The Softmax function is equivalent to a multinominal logistic transformation, which gives a probabilistic interpretation to the confidence values. The trained neural network has two hidden layers, and its structure is 22-28-26-4, due to the quantity of inputs. The first and last layers are the input and output layers, respectively. There are four output binary variables, which represent each of the semantic categories: closed_output (if it is the winner neuron, the concept is classified into the CLOSED semantic category), correct_output (CORRECT), incorrect_output (INCORRECT) and low_consumption_output (LOW CONSUMPTION). This ANN is used only to classify the different extracted concepts into a semantic category; it is not used during the analysis process. The ANN is used only in the analysis process when a new concept, which is not previously in the dictionary, appears. The generated dictionary is translated into COOL instance objects. The validation process was performed with 30% of the sample; additionally, the results were validated in the analysis process. It is not possible to check all of the concepts because such a check would require examining all of the consumers from the company databases. The new dictionary was tested on a well-known sample. The application of proposed neural network produced successful classification in approximately 27% of the extracted concepts. Thus, 49% of the extracted concepts was in the UNKNOWN semantic category, and they were discarded during the analysis process; however, the information about the frequency is updated in the dictionary at the end of the analysis process. The final classification summary is shown in Table 1.

The translation to COOL is performed by using objects that define the expression that is used to analyse the corresponding comment. The generated expressions are used in fixed rules that are always applied in the same way. Thus, the information is upgraded (using the previously described neural network) only on the new concepts, but the rules are always the same.

The description of this process is shown in the flow diagram of Fig. 4. This process is accomplished by means of engines that are implemented in SPSS Modeler, but it could also be performed by means of IBM Text Analytics because the results can be included in a stream of IBM SPSS Modeler.

These rules have different objectives:

– Gathering information about the consumer (the existence of battery of condensers, for example). In this case, the inspectors' commentaries contain one or more concepts that could be classified into the CB or LOCATION semantic category, or the extracted concept represents a telephone number or a contact name.
– Checking whether there has been some change in the economic activity. In this case, the inspectors' commentaries contain one or more concepts that could be classified into one or more of the 99 semantic categories about the economic activity, and usually the most recent one is selected.

Table 1
The final classification summary of concepts or multi-word relevant expression dictionary.

Sample description	Number of customers	Percentage
Initial Sample (economic activity related with general services)	52,098	100
Misinformation customers	2084	4
Analysed sample	50,014	96
Consumers selected by KBES with NTL	5136	9.9
Consumers selected by KBES for inspection	2403	4.6
Consumers finally sent to inspectors	136	0.26

– Characterisation of the consumption of the consumer. In this case, the inspectors' commentaries contain one or more concepts that can be classified into one or more of these semantic categories: CORRECT, INCORRECT, LOW CONSUMPTION or CLOSED.

4.4. Detection rules

The developed KBS uses the rules that have been implemented according to the knowledge of the inspectors and the rules that were generated by the methods of statistical analysis and text mining, which classify the consumers into different categories; a final conclusion is drawn as to whether the consumer has been inspected or not.

The KBS is composed of 177 rules, which can be classified according to their purpose or objective into the following groups:
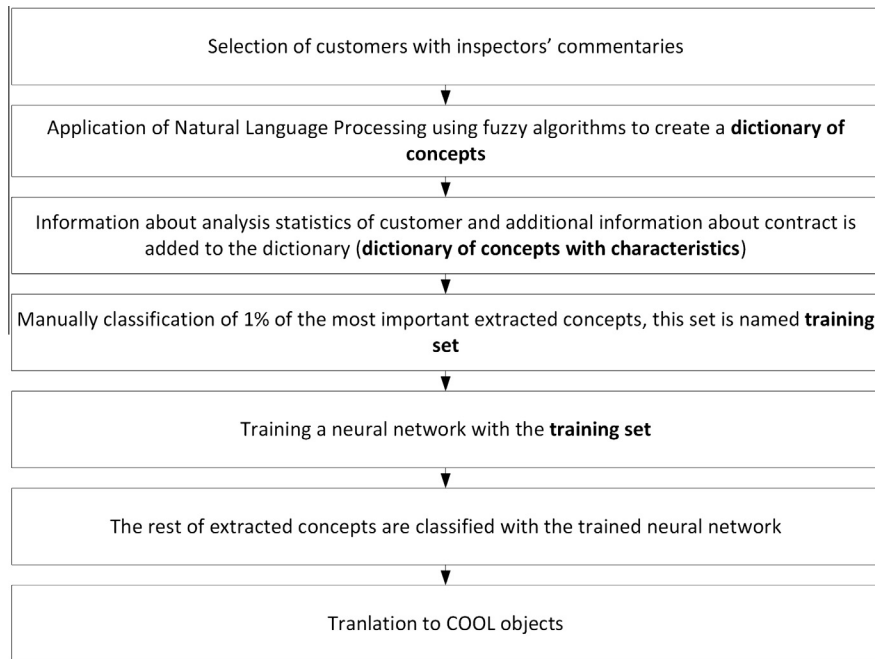
– 134 rules correspond to rules that are based on the knowledge of the inspectors. These rules are guided to inference different consumers' characteristics and to detect the different types of NTLs that can be identified from the data of the Endesa databases. There are mainly three types of problems: contract problems, measurement equipment problems and consumption problems. Contract problems can be solved without inspection (they are usually solved by a customer phone call). Measurement equipment problems are produced by inconsistencies in the information about the measurement equipment. Inspectors' commentaries may also suggest any type of problem. Consumption problems are typical problems that are related to the anomalous consumption of the client. All of these problems can identify or mask a possible NTL. For example, the following rule mainly checks whether the reactive energy consumption is higher than the active energy consumption. In addition, it checks whether the customer consumption is non-seasonal and whether the information about the measurement equipment is correct:

RULE NAME: reactive-higher-than-active
If (consumer has not seasonal consumption) and (consumer has at least the last three cycles of lectures with reactive energy greater than or equal to active energy) then inspection = YES and the activation of the rule is registered in the consumer analysis record endif

– 16 rules that are used to implement the revision on the comments of the inspectors (using knowledge that is generated by text mining). The objective of these rules is to interpret the knowledge about the consumer facilities according to the information that is stored in the databases. These rules use the information that is generated by text mining and the neural network module (previously described, in Section 4.3), which is translated into COOL object instances. In this way, the rules will

Fig. 4. The flow diagram that classifies the commentaries of the inspectors, which uses text mining and neural networks.

always be the same, but the information on text mining and neural networks can be updated. The following example is a rule that extracts information from commentaries on the historical information about the anomalies. This information could characterise the client consumption or provide new information.

RULE NAME: comment-h-anomaly
If in the commentaries of anomalies of each consumer (there are any identified economic activity and it is different from the economic activity specified in the contract and it is not previously identified) or (there are any classified concept on a known semantic category and it is not previously identified) then the economic activity or the semantic category identified is added to the consumer analysis record, a date is associated to it and the activation of the rule is added to the consumer analysis record endif

- 2 rules that are based on the use of the knowledge generated by the statistical analysis techniques. These rules are used to check whether the client consumption matches with the normal consumption pattern. These rules use the information that is generated by the statistical analysis module (previously described, in Section 4.2) that is translated into COOL object instances; in this way, the rules will always be the same, but the information on the statistical analysis can be updated.
- 20 rules that are used to preprocess the information. These rules check the information about the consumers and their installation, checking the consistency of the information.
- 4 rules that are used to debug the analysis process.
- 1 rule that is used to create the report on the analysis. This rule creates one report (plain-text format) that contains the following for each consumer: identification of the client, fired rules that identify an NTL, fired rules that discard the client, and the analysis results. This report is used by the IBM SPSS Modeler to make a new report that has graphical information, including the client consumption, contract characteristics, and the identification and location of the supply. This report is in HTML format. The final report shows several graphics about the consumption and readings. These graphics are shown in Figs.

5–7. This information is shown according to the inspectors' needs. It provides a quicker view of the status of the client facilities. An example of the report is shown in Fig. 8. This report was translated into English, but the translation is difficult to use because the Endesa staff wrote using informal language and they use imperative phrases and many abbreviations.

Each of these rules was verified and validated according to the knowledge of the inspectors. The rules were additionally validated using a well-known sample of consumers. In turn, the results shown in this article (next section) involved the validation step with real cases.

5. Experimental results

After performing all of the necessary tests to validate and verify the KBS, this KBS has been applied to a sample of 52,098 consumers in Catalonia (northwest of Spain) and corresponds to consumers who have economic activities or businesses that are related to the service economic sector. In this sample, the KBS did not classify 2084 consumers due to erroneous information. Thus, the final analysed and classified sample is composed of 50,014 consumers. A summary of this initial selection is shown in Table 2.

After the analysis process, the KBS selected approximately 10% of the supplies to be analysed. Overall, 5136 consumers were selected because they have different types of problems, but in fact, 2403 consumers needed only an inspection to confirm whether an NTL exists because the remaining consumers had problems that could be solved without an inspection.

The distribution company required the samples to not be very large, which enabled the management of the results and reduced the cost of the operations. In this way, 136 consumers were selected for inspection according to the number of rules that selected the consumers for inspection.

The inspectors of the distribution company inspected the set of 136 consumers. Finally, they performed 116 inspections. The remaining set of 20 consumers had failed visits because it was not possible to locate those consumers.
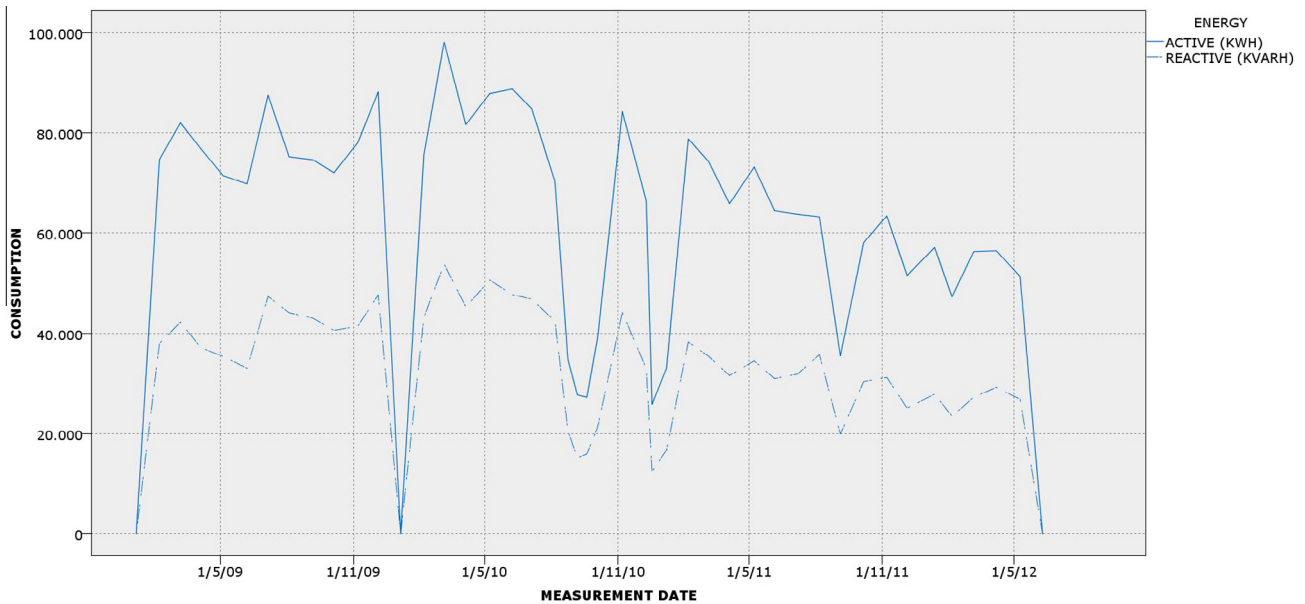
Fig. 5. Chart of energy active and reactive consumption over the past three years.
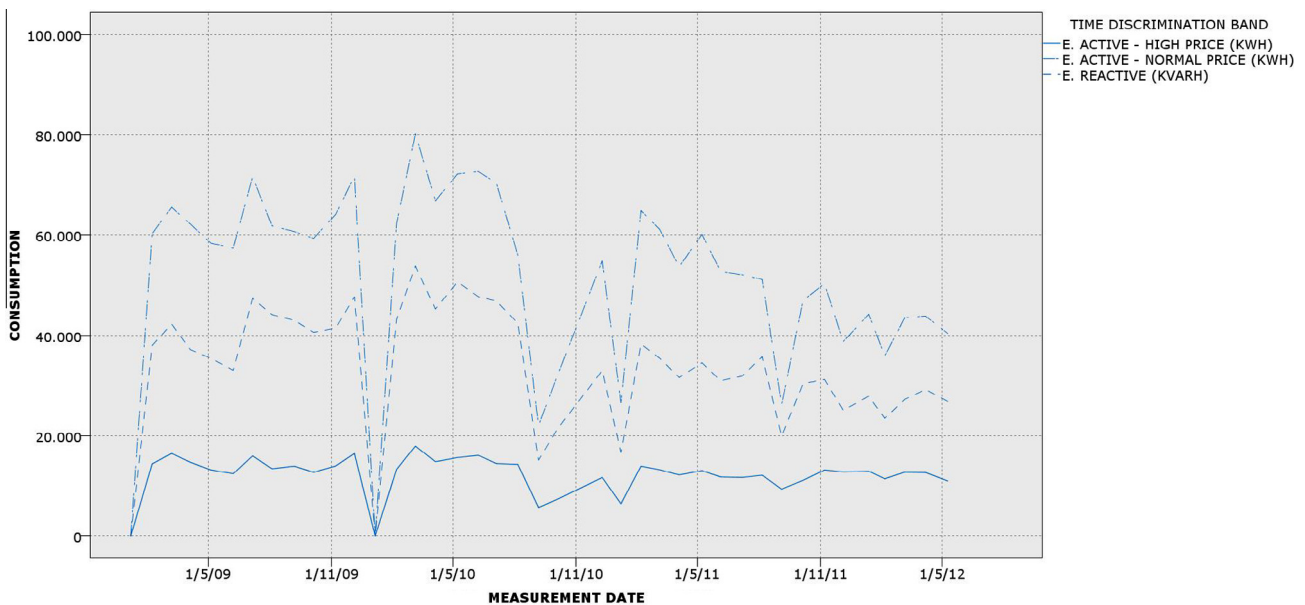

Fig. 6. Chart of energy per time discrimination bands over the past three years.

In the set of performed inspections (116 consumers), it is possible to highlight the following results (a summary of these results is shown in Fig. 9):

– 73 consumers did not present any NTL.
– 4 consumers were correct but had a different economic activity or business. There are several rules that apply different conditions according to the economic activity of the consumer; if this information is uninformed or mis-specified and there is no additional information in the inspectors' commentaries, then the result of the rule is not valid.
– 20 consumers who had anomalies but did not have demonstrable energy losses.
– 7 consumers who had anomalies with energy losses.
– 10 consumers with fraudulent activity and/or results.

– 2 consumers who withdrew from the supply. Usually, these consumers exhibit lower consumption.

A summary of these results is shown in Table 3.

In fact, there are 73 consumers who had correct consumption and 37 consumers who had NTLs (or situations that could mask an NTL). In this way, the success of the analysis was 33.6%; if the analysis is compared to all of the consumers, the success of the analysis would be 31.9%. The traditional methods of NTL detection in distribution companies have a success rate of approximately 10%.

The proposed KBS provides additional advantages that are related to the time taken for the analysis process. One inspector usually needs between 5 and 30 minutes to analyse all of the data of the consumer manually to determine whether there is an NTL.
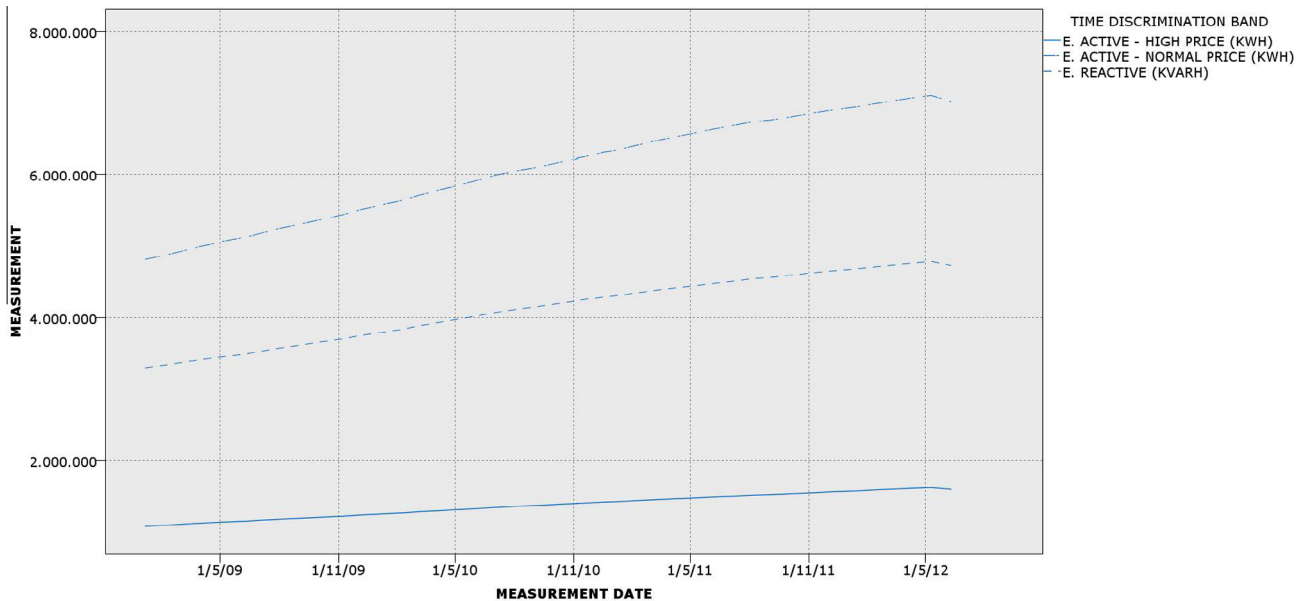
Fig. 7. Chart of lectures over the past three years.

Fig. 8. Translated report example.

Table 2
Summary of the initial sample.

Sample description	Number of customers	Percentage
Inspections carried out	116	100
Consumers who did not present any NTL	73	62.9
Consumers who are correct, but had different economic activity or business	4	3.4
Consumers with anomalies but without demonstrable energy losses	20	17.3
Consumers with anomalies and energy losses	7	6
Consumers with frauds	10	8.7
Consumers who withdraw the supply	2	1.7

The proposed KBS decreased the time that is taken to analyse the consumers. It is difficult to establish the time for the analysis because the system works in terms of two different types of software. When the system is working in IBM SPSS Modeler, the running of the analysis is faster because the software was designed to work with a large quantity of data (records), but in the case of CLIPS (and COOL), there are several memory restrictions, and it is necessary to use CLIPS environments. However, the analysis of 50,014 consumers was performed in 81 hours; thus, approximately 6 seconds was used for each consumer. If a person must perform this analysis without KBS help, it would take between 4167.8 to 25,007 hours/person; in this way, if one person must analyse 50,014 customers and must work for 8 hours per day, then this person would require from 2 to 12.4 years.

These tests were performed on a computer with Intel Core I5 650 (3.2 GHz S1156 4 MB), and 16 GB (DDR3 PC3-8500 1333 MHz) RAM was used to perform the analysis processes. This process can be simultaneously performed using several computers, and the time analysis can be reduced to less than 24 hours by using 4 computers.
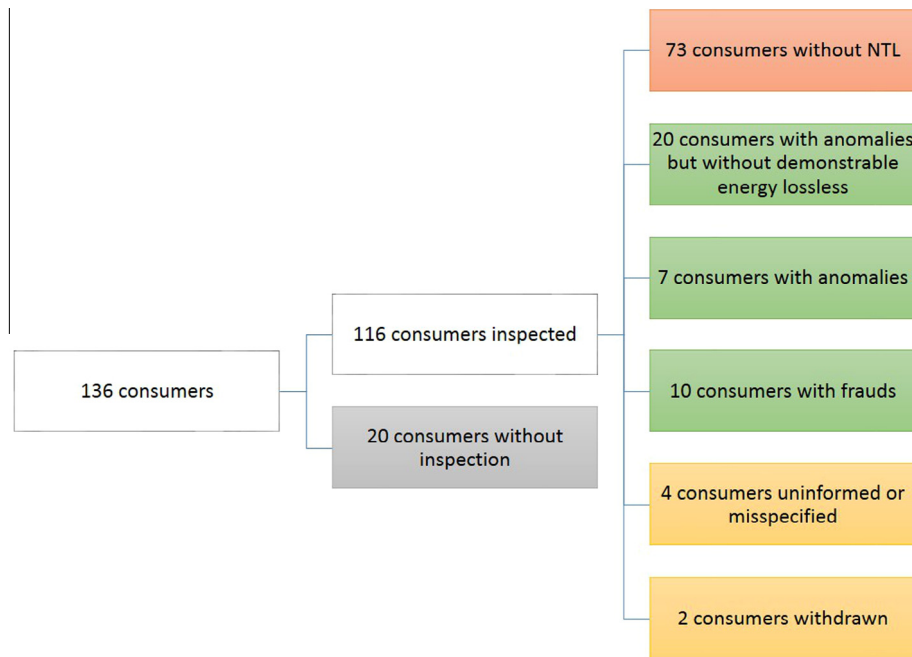
Fig. 9. Summary of the results.

Table 3
Summary of the selected sample.

Description	Percentage of classified concepts
The most frequent concepts which were manually classified	1
Remaining concepts: dates, numbers, and names	23
Extracted concepts which were automatically classified using the neural network	27
Extracted concepts automatically which were classified into UNKNOWN semantic category	49

6. Conclusions

The proposed KBS is composed of several sub-systems that provide additional knowledge about the customer. First, this KBS has 177 rules that represent the main knowledge base. Second, several statistical indicators are calculated using the information about customers with correct consumption. Third, a technique that is based on text mining and neural networks is applied to provide more information using the inspectors' commentaries. Finally, a report module is provided to show the information for the inspection process.

The proposed KBS can be used by any type of user and does not necessarily require advanced knowledge. Similarly, the system can be used as an educational tool because the KBS provides reports that explain the analysis results.

In addition, the time that is taken to analyse the consumers is highly reduced if it is compared with the time that is taken by the inspectors in the manual analysis. The proposed solution could also be produced using several computers, and the analysis process would be faster with the same human resources. The proposed KBS is in the test stage in Endesa, and this version has been used since the beginning of 2014.

Finally, the developed KBS has a success rate of 33.6%. This rate is higher than the rate obtained by using the traditional techniques of power distribution companies. Traditionally, the power distribution companies use a large number of consumers, who are selected

by means of a series of conditions and who are defined by the company; they have a success rate of approximately 10%.

The KBS uses the results of the analysis and inspections to provide feedback for subsequent analysis; in this way, the customers whose measurements were correct in this sample will be selected only if the consumption trend is different or has changed.

7. Future research directions

Currently, we are working on the implementation and testing of new rules that have been acquired in inspectors' interviews, and a new version is expected at the beginning of 2015.

In addition, we are increasing the efficiency of the analysis, and we could extend this analysis to other utilities, such as gas and water distribution; additionally, this analysis is to be tested with new smart meters and smart grid infrastructures, which provide additional information about consumers.

In future research that is related to this system, we could consider other technologies that are related to text mining, case-based reasoning (CBR), and association rules.

References

[1] C.R. Bayliss, B.J. Hardy, Smart grids, in: Transmission and Distribution Electrical Engineering, fourth ed., Newnes, Oxford, 2012, pp. 1059–1074 (Chapter 27).
[2] S.S.S.R. Depuru, L. Wang, V. Devabhaktuni, Electricity theft: overview, issues, prevention and a smart meter based approach to control theft, Energy Policy 39 (2) (Feb. 2011) 1007–1015.
[3] C. King, J. Strapp, Software infrastructure and the smart grid, in: Fereidoon P. Sioshansi (Ed.), Smart Grid, Academic Press, Boston, 2012, pp. 259–288 (Chapter 11).
[4] D.S. Markovic, D. Zivkovic, I. Branovic, R. Popovic, D. Cvetkovic, Smart power grid and cloud computing, Renew. Sustain. Energy Rev. 24 (August) (2013) 566–577.
[5] Expert Group 1: Functionalities of Smart Grids and Smart Meters, Final Deliverable, EU Commission Task Force for Smart Grids, December 2010.
[6] Position Paper on Smart Grids: An ERGEG Conclusions Paper, European Regulators Group for Electricity & Gas, E10-EQS-38-05, June 2010.
[7] Strategic Deployment Document for Europe's Electricity Networks of the Future, European Technology Platform SmartGrids, April 2010.
[8] F. Bouhafs, M. Mackay, M. Merabti, Links to the future: communication requirements and challenges in the smart grid, IEEE Power Energy Mag. 10 (1) (2012) 24–32.

  9 E.D. Knapp, R. Samani, What is the smart grid?, in: E.D. Knapp, R. Samani (Eds.), Applied Cyber Security and the Smart Grid, Syngress, Boston, 2013, pp. 1–15 (Chapter 1).

[10] S. Luthra, S. Kumar, R. Kharb, M.F. Ansari, S.L. Shimmi, Adoption of smart grid technologies: an analysis of interactions among barriers, Renew. Sustain. Energy Rev. 33 (May) (2014) 554–565.

[11] M. Chebbo, EU SmartGrids framework 'electricity networks of the future 2020 and beyond', in: IEEE Power Engineering Society General Meeting, 2007, 2007, pp. 1–8.

[12] V. Giordano, A. Meletiou, C. Felix Covrig, M. Ardelean, G. Fulli, M. Sánchez Jiménez, C. Filiou, Smart Grid Projects in Europe: Lessons Learned and Current Developments, European Commission, Luxembourg, EUR 25815 EN, 2013.

[13] The Global Smart Grid Federation Report, 2012.

[14] P. Lewis, Smart Grid 2013 Global Impact Report, Ventix, VaasaETT Global Energy Think Tank, 2013.

[15] M.P. McHenry, Technical and governance considerations for advanced metering infrastructure/smart meters: technology, security, uncertainty, costs, benefits, and risks, Energy Policy 59 (Aug. 2013) 834–842.

[16] C. Selvam, K. Srinivas, G.S. Ayyappan, M. Venkatachala Sarma, Advanced metering infrastructure for smart grid applications, in: 2012 International Conference on Recent Trends In Information Technology (ICRTIT), 2012, pp. 145–150.

[17] L. Dan, H. Bo, Advanced metering standard infrastructure for smart grid, in: 2012 China International Conference on Electricity Distribution (CICED), 2012, pp. 1–4.

[18] M. Naglic, A. Souvent, Concept of SmartHome and SmartGrids integration, in: 2013 4th International Youth Conference on Energy (IYCE), 2013, pp. 1–5.

[19] Z. Luhua, Y. Zhonglin, W. Sitong, Y. Ruiming, Z. Hui, Y. Qingduo, Effects of advanced metering infrastructure (AMI) on relations of power supply and application in smart grid, in: 2010 China International Conference on Electricity Distribution (CICED), 2010, pp. 1–5.

[20] P. Siano, Demand response and smart grids—a survey, Renew. Sustain. Energy Rev. 30 (February) (2014) 461–478.

[21] E. Valigi, E. Di Marino, Networks optimization with advanced meter infrastructure and smart meters, in: 20th International Conference and Exhibition on Electricity Distribution – Part 1, 2009, CIRED 2009, 2009, pp. 1–4.

[22] Z. Vale, H. Morais, P. Faria, C. Ramos, Distribution system operation supported by contextual energy resource management based on intelligent SCADA, Renew. Energy 52 (Apr. 2013) 143–153.

[23] M.L. Di Silvestre, E. Riva Sanseverino, Modelling energy storage systems using Fourier analysis: an application for smart grids optimal management, Appl. Soft Comput. Part C 14 (2014) 469–481.

[24] F. Javed, N. Arshad, F. Wallin, I. Vassileva, E. Dahlquist, Forecasting for demand response in smart grids: an analysis on use of anthropologic and structural data and short term multiple loads forecasting, Appl. Energy 96 (Aug. 2012) 150–160.

[25] D.-H. Yang, S. Kim, C. Nam, J.-W. Min, Developing a decision model for business process outsourcing, Comput. Oper. Res. 34 (12) (2007) 3769–3778.

[26] J.R. Aguero, Improving the efficiency of power distribution systems through technical and non-technical losses reduction, in: Transmission and Distribution Conference and Exposition (T D), 2012 IEEE PES, 2012, pp. 1–8.

[27] V. Paruchuri, S. Dubey, An approach to determine non-technical energy losses in India, in: 2012 14th International Conference on Advanced Communication Technology (ICACT), 2012, pp. 111–115.

[28] J.M.R. Iglesias, Follow-up and preventive control of non-technical losses of energy in C.A. Electricidad de Valencia, in: Transmission Distribution Conference and Exposition: Latin America, 2006, TDC'06. IEEE/PES, 2006, pp. 1–5.

[29] R. Alves, P. Casanova, E. Quirogas, O. Ravelo, W. Gimenez, Reduction of non-technical losses by modernization and updating of measurement systems, in: Transmission Distribution Conference and Exposition: Latin America, 2006, TDC'06, IEEE/PES, 2006, pp. 1–5.

[30] A.H. Nizar, Z.-Y. Dong, P. Zhang, Detection rules for non technical losses analysis in power utilities, in: 2008 IEEE Power and Energy Society General

Meeting – Conversion and Delivery of Electrical Energy in the 21st Century, 2008, pp. 1–8.

[31] S.S.S.R. Depuru, L. Wang, V. Devabhaktuni, Support vector machine based data classification for detection of electricity theft, in: Power Systems Conference and Exposition (PSCE), 2011 IEEE/PES, 2011, pp. 1–8.

[32] C.C.O. Ramos, A.N. Souza, R.Y.M. Nakamura, J.P. Papa, Electrical consumers data clustering through Optimum-Path Forest, in: 2011 16th International Conference on Intelligent System Application to Power Systems (ISAP), 2011, pp. 1–4.

33 J.J. Bann, B.S. Baer, 33 – Expert systems in power systems control, in: Cornelius T. Leondes (Ed.), Knowledge-Based Systems, Academic Press, San Diego, 2000, pp. 1061–1108.

[34] Y.O. Yang, S.H. Chang, A diagnostic expert system for the nuclear power plant based on the hybrid knowledge approach, IEEE Trans. Nucl. Sci. 36 (6) (1989) 2450–2458.

[35] Q. Zhang, A frequency and knowledge tree/causality diagram based expert system approach for fault diagnosis, Reliab. Eng. Syst. Saf. 43 (1) (1994) 17–28.

[36] M. Ross, R. Hidalgo, C. Abbey, G. Joós, Energy storage system scheduling for an isolated microgrid, IET Renew. Power Gener. 5 (2) (2011) 117–123.

[37] D. Yuehui, Y. Yihan, A knowledge- and network-based cognition model expert system of automatic switching sequences for substations and plants, in: 1998 International Conference on Power System Technology, 1998, Proceedings, POWERCON'98, vol. 2, 1998, pp. 982–987.

[38] B. Picard, F.D. Galiana, D. Mcgillis, A knowledge-based system for the structural design of high-voltage lines, in: 1999 IEEE Canadian Conference on Electrical and Computer Engineering, vol. 3, 1999, pp. 1229–1233.

[39] M.S. Kandil, S.M. El-Debeiky, N.E. Hasanien, Long-term load forecasting for fast developing utility using a knowledge-based expert system, IEEE Trans. Power Syst. 17 (2) (2002) 491–496.

[40] V.V.K. Reddy, M. Sydulu, A Heuristic-Expert based Approach for Reconfiguration of Distribution Systems, Presented at the IEEE Power Engineering Society General Meeting, 2007, pp. 1–4.

[41] A.Y. Abdelaziz, S.F. Mekhamer, M.H. Nada, A fuzzy expert system for loss reduction and voltage control in radial distribution systems, Electr. Power Syst. Res. 80 (8) (2010) 893–897.

[42] L. Šubelj, Š. Furlan, M. Bajec, An expert system for detecting automobile insurance fraud using social network analysis, Expert Syst. Appl. 38 (1) (2011) 1039–1052.

[43] C.S. Hilas, Designing an expert system for fraud detection in private telecommunications networks, Expert Syst. Appl. 36 (9) (2009) 11559–11569.

[44] K.J. Leonard, Detecting credit card fraud using expert systems, Comput. Ind. Eng. 25 (1–4) (1993) 103–106.

[45] K.J. Leonard, The development of a rule based expert system model for fraud alert in consumer credit, Eur. J. Oper. Res. 80 (2) (1995) 350–356.

[46] J.I. Guerrero, C. León, S. Member, F. Biscarri, Í. Monedero, J. Biscarri, R. Millán, A real application on non-technical losses detection: the MIDAS Project, in: Proceedings of The 2011 International Conference on Data Mining (DMIN'11), CSREA Press, Las Vegas, Nevada, 2011.

[47] C. León, F. Biscarri, I. Monedero, J.I. Guerrero, J. Biscarri, R. Millán, Integrated expert system applied to the analysis of non-technical losses in power utilities, Expert Syst. Appl. 38 (8) (2011) 10274–10285.

[48] I. Monedero, F. Biscarri, C. León, J.I. Guerrero, J. Biscarri, R. Millán, Detection of frauds and other non-technical losses in a power utility using Pearson coefficient, Bayesian networks and decision trees, Int. J. Electr. Power Energy Syst. 34 (1) (2012) 90–98.

[49] C. Leon, F. Biscarri, I. Monedero, J.I. Guerrero, J. Biscarri, R. Millan, Variability and trend-based generalized rule induction model to NTL detection in power companies, IEEE Trans. Power Syst. 26 (4) (2011) 1798–1807.

[50] J.I. Guerrero, C. Leon, F. Biscarri, I. Monedero, J. Biscarri, R. Millán, Increasing the efficiency in Non-Technical Losses detection in utility companies, in: MELECON 2010–2010 15th IEEE Mediterranean Electrotechnical Conference, 2010, pp. 136–141.

[51] S. Linckels, C. Meinel, Natural language processing, in: E-Librarian Service, Springer, Berlin Heidelberg, 2011, pp. 61–79.