Knowledge-Based Systems 71 (2014) 376–388

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

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

A R T I C L E I N F O

Article history:
Received 22 December 2013
Received in revised form 12 August 2014
Accepted 13 August 2014
Available online 23 August 2014

Keywords:
Expert system
Power distribution
Non-technical losses
Neural network
Text mining

A B S T R A C T

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.

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).

http://dx.doi.org/10.1016/j.knosys.2014.08.014
0950-7051/

– 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