Chapter 47
Detection of Non-Technical Losses: The Project MIDAS

Juan I. Guerrero
Universidad de Sevilla, Spain

Jesús Biscarri
Universidad de Sevilla, Spain

Íñigo Monedero
Universidad de Sevilla, Spain

Rocío Millán
Universidad de Sevilla, Spain

Félix Biscarri
Universidad de Sevilla, Spain

Carlos León
Universidad de Sevilla, Spain

ABSTRACT

The MIDAS project began in 2006 as collaboration between Endesa, Sadiel, and the University of Seville. The objective of the MIDAS project is the detection of Non-Technical Losses (NTLs) on power utilities. The NTLs represent the non-billed energy due to faults or illegal manipulations in clients' facilities. Initially, research lines study the application of techniques of data mining and neural networks. After several researches, the studies are expanded to other research fields: expert systems, text mining, statistical techniques, pattern recognition, etc. These techniques have provided an automated system for detection of NTLs on company databases. This system is in the test phase, and it is applied in real cases in company databases.

INTRODUCTION

The main objective of data mining techniques is the evaluation of data sets to discover relationships in information. These relationships may identify anomalous patterns or patterns of frauds. Fraud detection is a very important problem in telecommunication, financial and utility companies. Currently data mining is one of the most important techniques which are applied to solve these types of problems, joined with: rough sets, neural networks, time series, support vector machines, etc. There are a lot of references about the detection of abnormalities or frauds in a set of data.

The increase of storage capacity and the process capacity allow one to manage large databases. Data

DOI: 10.4018/978-1-4666-9562-7.ch047

mining provides a set of techniques of artificial intelligence which can be used to increase the efficiency of data mining methods.

The utility companies have large databases which support the management processes. In addition, these companies invest their effort in maintenance of infrastructure and anomaly detection. These anomalies are frauds in telecommunication and financial sectors; breakdown or fraud in power, water or gas sectors; etc.

The non-technical losses (NTLs) in power utilities are defined as any consumed energy or service which is not billed because of measurement equipment failure or ill-intentioned and fraudulent manipulation of said equipment. This paper describes advances developed for the MIDAS project. The paper proposes a framework to analyze all information available about customers. This framework uses: data mining, text mining, expert systems, statistical techniques, regression techniques, etc. The proposed framework is actually in the testing phase. It is the main result of the MIDAS Project a collaborative project between the Endesa Company, Ayesa and the University of Seville.

In this paper, a description of the framework is made, following these steps:

- Review of current state about the anomaly detection and NTLs detection. Additionally, the Endesa utility company is described.
- The MIDAS project is explained.
- Each module is described.
- Finally, the conclusions are presented.

REVIEW OF THE CURRENT STATE

Bibliographical Review

The Non-Technical Losses (NTLs) were increasingly regarded as a cause of concern in distribution utility companies. There exists several causes of NTLs and they can affect quality of supply, electrical load on the generating station and tariff imposed on electricity consumed by genuine customers. (Depuru, Lingfeng Wang, Devabhaktuni, & Gudi, 2010) discusses various factors those influence the consumer to make an attempt to steal electricity. There are a lot of methods for detection NTL. The distribution utility companies are interested in analysis of NTLs for detection, location and classification of NTLs, with the objective of reducing them. There are many ways to perform these processes, and can be taken as reference similar methods used for fraud detection in telecommunications, finance, etc. (Yufeng Kou, Chang-Tien Lu, Sirwongwattana, & Yo-Ping Huang, 2004) and (Weatherford, 2002) show different techniques related with data mining for fraud detection, including the most interesting parameters for using with them.

Financial Sector

In the financial sector there are a lot of references with the use of data mining and computational intelligence in the fraud detection. Noteworthy is the use of these techniques in credit card fraud detection. In the nineties, (Ghosh & Reilly, 1994), (Fanning, Cogger, & Srivastava, 1995), (Aleskerov, Freisleben, & Rao, 1997) and (Dorronsoro, Ginel, Sgnchez, & Cruz, 1997) used neural networks to detect ones. Some authors publish researches with other techniques: intelligent hybrid system (Hambaba, 1996), neural networks compared to statistical techniques (Richardson, 1997), neural data mining (Brause, Langsdorf, & Hepp, 1999), distributed data mining (Chan, Fan, Prodromidis, & Stolfo, 1999), etc. In addition, other techniques are used latter, for example: Genetic Algorithm (Özçelik, Işik, Duman, & Çevik, 2010), neural networks and logistic regression (Y. Sahin & Duman, 2011), time series (Seyedhossein & Hashemi, 2010), decision trees (Yusuf Sahin, Bulkan, & Duman, 2013), etc.

In financial sector, there are other areas of interest, for example, based in the theory of Rough