UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR

PROYECTO FIN DE CARRERA

DECISION SUPPORT SYSTEM: KNOWLEDGE CAPTURE AND SHARING FOR TELECOM NETWORK MANAGEMENT

Ingeniería de Telecomunicación

Berta Lorenzo Espinar

Noviembre 2014

DECISION SUPPORT SYSTEM: KNOWLEDGE CAPTURE AND SHARING FOR TELECOM NETWORK MANAGEMENT

Berta Lorenzo Espinar

TUTOR: Shahina Begum

PONENTE: Daniel Ramos Castro

Escuela Politécnica Superior
Universidad Autónoma de Madrid
Noviembre de 2014

Abstract

This thesis work presents analysis and processing of the data from a European Distribution System Operator (DSO) in order to develop a decision support system which could help field technicians during the maintenance and operation of the electricity network. The proposed system is based on an artificial intelligence method called Case-Based Reasoning (CBR) which uses previous experiences to solve a new problem case. The four steps in this algorithm are: retrieval, reuse, revise and retain.

The main stages describe in this report are problem domain analysis, data analysis, knowledge acquisition, feature extraction, case representation and development of a case library for the CBR system. However, implementation of the CBR cycle has not been included in this work; only an initial retrieval approach has been discussed using the tool *myCBR*. The initial result shows that the selected features could help to develop a case-library for building a system for fault diagnosis in the domain.

This thesis work is a part of a project, where the final goal is to automate the fault analysis process for the technician using analytics and knowledge reasoning in a Distribution System Operator. This project started in May 2014 at Ericsson Research offices in Kista, Stockholm.

Key words: Decision Support System, Case-Based Reasoning, Telecom Network, Case Library, Features

Resumen

Este Proyecto de Fin de Carrera presenta el análisis y procesamiento de los datos de una red de distribución de energía eléctrica (DSO) con el fin de desarrollar un sistema de apoyo a decisiones que podría ayudar a los técnicos durante el mantenimiento y operación de la red eléctrica. El sistema propuesto se basa en un método de inteligencia artificial llamado Razonamiento Basado en Casos (CBR), que utiliza las experiencias previas para resolver los nuevos problemas. El algoritmo CBR se compone de cuatro etapas: la recuperación de casos anteriores, la reutilización de la solución de los casos obtenidos, la revisión de dicha solución del nuevo caso, y, por último, el almacenamiento del nuevo problema y su solución en la base de datos del sistema.

Las principales tareas que se describen en esta memoria son el análisis del problema y de la red eléctrica, el análisis de datos, la adquisición de información, la extracción de características, la representación de los casos y el desarrollo de una base de datos específica para el sistema CBR propuesto. Sin embargo, la aplicación de las cuatro etapas de CBR previamente mencionadas no se han incluido en este trabajo; sólo un planteamiento inicial de la fase de recuperación, que ha sido propuesto utilizando la herramienta llamada *myCBR*.

El resultado final de este proyecto es la elaboración de una base de datos con los problemas típicos en los medidores de la red eléctrica y la definición de un sistema basado en la técnica CBR que pueda ayudar durante el diagnóstico y resolución de problemas en dicha red.

Esta memoria forma parte del proyecto que se está llevando a cabo en las oficinas de Ericsson Research AB (Suecia). El objetivo final de este proyecto es automatizar el proceso de análisis de fallos en la red eléctrica, utilizando técnicas de análisis y razonamiento basadas en el conocimiento previo. Este proyecto se inició en mayo de 2014, a las oficinas de Ericsson Research en Kista, Estocolmo.

Palabras clave: sistema de ayuda a decisiones, razonamiento basado en casos, red, base de datos, características

Acknowledgments

First, I would like to thank my advisors in Sweden, Shahina Begum and Peter Funk, for the opportunity of performing this Master Thesis at Mälardalen University in Västerås (Sweden) in collaboration with Ericsson Research AB. Thanks also to my advisors at Ericsson: Hongxin Liang, Gábor Stikkel and Leonid Mokrushin. Also, to all the people who have helped me during the past months.

I would also like to thank Daniel Ramos for assisting me with the Master Thesis at the Universidad Autónoma de Madrid, and my teachers to help all of us build up a future in the world of Telcos.

Special mention to my fellows at the University for the special moments in this important stage of our lives. I have shared with them the laboratories or the library days, but also breaks, parties and trips during these years. Because this friendship should last many years, thanks to the Toligos (Eva, Fatima, Helia, James, Maria, Ricardo, Rodrigo, Sandra and Sara) and my other friends.

I cannot forget all the people I have met during the Erasmus and have been my family. Ana, my wife, for all those funny moments that we only understood and for being my perfect travelling partner. Sergio, for supporting me when the project didn't go ahead and don't leave me alone at any time. And to the others, Borja, Ángel, Sama, Richi, Sejo and Carlos, thanks for these months in Sweden, for the good times and all the beers have fallen when we were "de tranquis".

Finally, I would like to thank my parents, who have supported me over my entire life and who have done everything necessary for their daughters. My father, Bernardo, has taught me what the effort is and how hard it is to get things. My mother, Esther, has helped every day and has educated myself to see what the real life is. To my sister, Bea, she is always trying to make the others smile and she is always there for me. Thanks also to my grandparents, uncles and cousins, for your love.

Thank you so much. Tack så mycket.

Berta
November 2014

Agradecimientos

En primer lugar me gustaría agradecer a mis tutores en Suecia, Shahina Begum y Peter Funk, por ofrecerme la oportunidad de realizar este Proyecto de Fin de Carrera en la Mälardalen University en Västerås (Suecia) en colaboración con Ericsson Research AB. Gracias también a mis tutores en Ericsson: Hongxin Liang, Gábor Stikkel y Leonid Mokrushin. Así mismo, a todas las personas que me han ayudado durante los últimos meses.

Me gustaría también agradecer, en primer lugar, a Daniel Ramos por aceptar ser mi ponente en la Universidad Autónoma de Madrid, y a todos los profesores de la carrera por ayudarnos a labrarnos un futuro en el mundo de las Telecos.

Se merecen una mención especial mis compañeros de carrera, por todos los momentos vividos durante esta etapa de nuestras vidas. Con ellos he compartido tanto los interminables ratos en los laboratorios o los días de biblioteca, como los descansos entre clases, las fiestas y los viajes de los últimos años. Por vuestra ayuda durante toda la carrera y porque esta amistad dure muchos años más, gracias a los Toligos (Eva, Fátima, Helia, Jaime, Maria, Ricardo, Rodrigo, Sandra y Sara) y al resto de compañeros de la EPS.

No me quiero olvidar tampoco de la gente que he conocido durante el Erasmus y que ha sido mi familia cuando más lo he necesitado. Ana, mi wife, por todos esos ratos de risas que sólo nosotras entendíamos y por ser mi compañera de viajes perfecta. Sergio, por apoyarme cuando el proyecto no salía adelante y no dejarme sola en ningún momento. Y a los demás, qué decir, Borja, Ángel, Sama, Richi, Sejo y Carlos, gracias por estos meses en Suecia, por los buenos ratos y por todas las cervezas que han caído cuando íbamos "de tranquis".

Dejo para el final a mis padres, quienes más me han apoyado durante toda mi vida y quienes han hecho todo lo necesario para que sus hijas salieran adelante. Sin ellos nunca hubiera podido llegar hasta aquí. Mi padre, Bernardo, me ha enseñado qué es el esfuerzo y lo difícil que es conseguir las cosas. Mi madre, Esther, por ayudarme a conseguir todo lo que me proponía y por educarme para que viera la vida real y tuviera los pies en la tierra. Por otro lado, gracias a mi hermana, Bea, que siempre tiene la energía suficiente para sacarte una sonrisa y por estar siempre ahí. Gracias también a mis abuelos, tíos y primas, por vuestro cariño.

Muchas gracias a todos.

Berta
Noviembre 2014

Contents

List of Figures

List of Tables

Glossary

AI	Artificial Intelligence
CBR	Case-Based Reasoning
DSS	Decision Support System
DSO	Distribution System Operator
ES	Expert System
IDSS	Intelligence Decision Support Systems
KB-CBR	Knowledge Based – Case-Based Reasoning
KM	Knowledge Management
NOC	Network Operation Center

1. INTRODUCTION

Today, the telecommunication industry is an extended and accessible market to companies in all the possible fields of study. Around the world there are many companies which would like to expand their systems in order to get the best position in the market they belong to.

Due to rapid advancement and competitive market companies need to update their software and, at this moment, adding computer-aided diagnostic system is one of the useful processes in order to save time and money. When a company decides to expand its business and grow in order to get more benefits, it may need to automate and update its software and adapt into the new technologies.

Artificial Intelligence (AI) methods and techniques can help in solving problems and support in making decisions. AI is a term that attends to understand human intelligence. Therefore, to develop systems based on human's behavior, first of all, it is needed to study human behavior and, after understanding the intelligence behavior of human, it is possible to make computers work in a similar way. Thus, AI systems are based on these studies. In other words, computers should possess the following capabilities: Natural language processing, knowledge representation, automated reasoning and machine learning.

This thesis is focusing on designing a Decision Support System (DSS) applying AI methods and techniques in maintenance and operation of electricity network. The DSS should help manager or field engineer to solve the possible problems in the network and decide a solution in the shortest time. The solution to solve the current problem will be suggested based on previous experience and human reasoning. In AI, systems using past knowledge and reasoning by analogy can be implemented using Case-Based Reasoning (CBR) algorithm [1] [2] [3].

1.1. Motivation

In telecommunication companies, Network operation centers help companies to administrate, control and monitor their networks. Companies with a huge number of customers need tools to control and detect problems in their networks. Applications and tools using today are supporting companies in some tasks, however, when they are planning to expand their networks it is essential to update and automate the corresponding functionalities/systems.

A field engineer should solve the problems in the network with information stored in the network operations centers but often it is not as easier as they can imagine. Frequently, they need more specific information and knowledge to find the real cause of the problem and define the best solution in this case. DSS could be a useful during the colleting process. Nowadays, research focuses on diagnostic and decision system using reasoning and rationality algorithms according to human behavior. Therefore a computer-based diagnostic applications enable real-time monitoring of a network and detect changes during operation would be beneficial in this domain. The systems should be able to control and detect events in real time as well as help administrator to solve problems in the shortest time to avoid bigger problems and large losses to the companies. Detecting alarms in the networks, getting information from sensor and nodes and proposing valuable solutions are important in diagnostic systems.

In this thesis work, the proposal is to design a decision support system based on case-based reasoning methods which stores all the relevant information of the electricity network and helps during problems resolution. This system will support field engineers and the network operations center using knowledge and experienced from the past.

Case-Based Reasoning (CBR) uses previous experience to solve a new problem. Therefore, it is needed to figure out the analogy between a current problem and the past problems to reuse past solution and give a correct solution to our system. In this method, there are four important steps to follow: compare problems and choose the closest one, get the information and solution of the previous problem, readapt the solution and probe and add the new experience as a machine learning method.

The main focus of this thesis work is to analyze the data from the electricity network and define key features to represent a problem case in order to build a case library. An efficient technique to extract key features is also essential to define similarity between the problems.

1.2. Background

Electricity networks comprise both transmission infrastructure and distribution infrastructure. Distribution is then about distributing and delivering this power to final customers, via smaller power lines at medium and/or lower voltage which form the distribution networks. Distribution System Operators (DSO) has the responsibility to deliver energy of suppliers to end-users and to maintain the distribution networks. They are hence particularly important in ensuring faultless delivery of electrical power to the end users.

In order to improve the quality of the customer service, the DSO has define two different goals: first of all, the renovation of the distribution network and the installation of a new metering system where all the meters will be monitored and controlled in order to offer a better energy price to the end customers. Smart meters will help the DSO in their market facilitation task. The smart metering system should be determined by how it can improve customer service and retail functioning while at the same time keeping costs down. On the other hand, the time needed for solving customer calls when a network problem occurs should be reduced. Problem solving times in distribution network are sometimes too long and thus cost too much; for this reason, it is needed a decision support system able to give answers in the shortest time.

1.3. Objectives

The main goal of this thesis work is to analyze the data from the electricity network operations system to propone a simple decision support system able to solve the meters problems as soon as possible. This decision support system will be based on case-based reasoning method.

For this purpose, there are two different objectives to describe.

1) Analyze the data from collected from the system and investigate different feature extraction techniques. Here, whenever a problem is reported in the system, a new case is created. So, the main idea is to propose a solution for the new case based on previous

experiences (past experienced, previous problems, similar cases) using the data collected from different parts of the system

2) Develop a prototype of the case-based reasoning system that helps the electricity network operations system in the tasks of analyzing problems, solving incidences and giving solutions in shortest time to save money and time.

1.4. Methodology

The following tasks are proposed to perform the project:

1. Understand the electricity network of the current system
2. Analyze the existing data and requirements of the current system in order to define the functionalities of the new system
3. Study different methods of developing a Decision Support System
4. Propose an AI method: Case-Based Reasoning
5. Develop a prototype of Decision Support System
6. Evaluation of the proposed system and conclusions

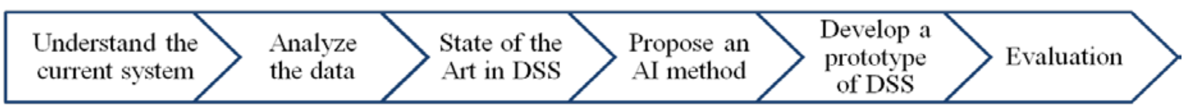

Figure 1-1: Methodology

1.5. Structure

Chapter 1 presents the motivation and objectives of this thesis work.

Chapter 2 describes an overview of the application domain and the method used in this thesis work. It describes case-based reasoning and helps to understand systems based on this method. This chapter also includes limitations and advantages of the methods.

Chapter 3 describes the use of CBR techniques in different areas and the research advances that uses this technique in our days.

Chapter 4 explains the background of the problem and describes the system that we are based on in this thesis work.

Chapter 5 describes the data analysis process and also the different problems and solutions of the electricity network.

Chapter 6 presents the steps followed to perform the case library for this thesis work and presents the proposed system.

Chapter 7 explains the evaluation of the CBR system and the results.

Chapter 8 presents the conclusion of the thesis work and introduces the future work.

2. BACKGROUND

This chapter provides a description about the concepts that the reader needs to know to understand the thesis work. First of all, the reader can find the definition of Decision Support Systems (DSS) and the different interpretations of it over the last years. The following explains the evolution of DSS and the methods used to develop these systems. Finally, related work discussing the use of CBR techniques in DSS.

2.1. Decision Support Systems

Now-a-days, control and management of networks operation centers using decision support systems is very common practice in companies of different domains. When making decision, information is required. A system to manage all the information and to help to make a good decision will help in management practice within companies.

Since the earlier 1970s, the concept of Decision Support System (DSS) has different understanding. DSS technology and applications have evolved significantly during the last 40 years. The original definition of DSS has introduced by Gorry and Morton in 1971, they introduced DSS as “interactive computer-based systems, which help decision makers utilize data and models to solve unstructured problems”. [8]

In 1978, DSS has been defined as a system that combines the intellectual resources of individuals with the capabilities of the computer to improve the quality of decisions.

It is a computer based support system for management decision makers who deal with semi-structured problems. [9]

According to Bonczek & Whinston [10] definition, DSS is a computer-based system used as a decision-maker that should contain three interacting components: a language system to provide communication in the DSS, a knowledge system in the problem domain and a problem-processing system. When experiencing a problem in a specific problem domain, a decision-maker uses information from a specific domain expert and knowledge from this problem domain, hence, a good communication between them it also needed.

The latest definitions of DSS describe it as a computerized information system that collects, organizes and analyzes data to supports business and organizational applications making decisions. A well-designed DSS helps managers to identify and solve problems and assists decision makers. These systems usually use an interactive software-based system that compiles all the information from raw data, documents, previous cases, personal knowledge and business models. [12]

When building a DSS some conceptions should be follow. First of all, a DSS is defined in terms of the structure of tasks it addresses, because it requires a distinctive design strategy based on evolution and “middle-out” techniques.

On the other hand, a DSS supports the cognitive processes of individual decision makers; decision research provides descriptive insights into management problem-solving and normative theories for defining how to improve its effectiveness. DSS are designed to combine human making-decision process and computer analyzes systems. Solutions could be different depending on the person who makes the decision due to several aspects as position, experience and knowledge. [7]

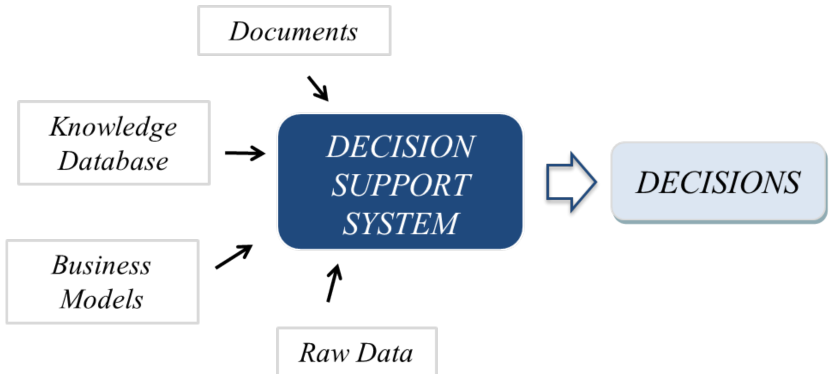

Figure 2-1: Decision support system

Decision makers also use information from expert domain and experience. An expert is who has specific knowledge in the problem domain and who has acquired extensive experience from training, reading out and solving similar problems in a specific domain. When solving complex problems, advice and knowledge from experts is essential to make better and faster decision than non-experts systems.

In Intelligence Artificial (AI) field, DSS and Experts systems (ES) are included in several tools to help and assist people in specialized situations. Both system will provide advice and help for making decision, but an ES will give better solutions when expertise is required. Hence, an ES application makes inferences and arrives at conclusions while a DSS provides an environment to assist a user to reach conclusions.

AI attempts to incorporate human's intelligence behavior and computer programs in a same system. Intelligent behaviors include cognitive skills like thinking, problem solving, learning and understanding that now could be included in several knowledge-based systems. Therefore, when knowledge and experience is needed, especially in networks operation centers where reducing costs and time during solving problems tasks is one of the most important goals, a Knowledge-based Decision support systems (KB-DSS) is important.

2.2. Knowledge-based Decision Support Systems

Knowledge-based Decision support systems (KB-DSS), also called Intelligent Decision Support Systems (IDSS), are included in the applications of AI. KB-DSS integrate traditional DSS and the advances of ES. Traditionally DSS constitute data management, modelling and decision methodology, but incorporating advances of ES, the KB-DSS include reasoning and explanation capabilities. [16]

KB-DSS are used to solve problems using previous knowledge, experience and other cognitive skills like learning or understanding in computer-based systems. The computer-based systems handled engineer experience about the problem domain and techniques that lead how to use the knowledge to solve problems in the same way as an expert would do.

The architecture of KB-DSS includes two main sub-systems: the knowledge base, which contains all the knowledge required to solve problems in the specific domains, and the interference engine, which represents the technique that manipulates, uses and controls the knowledge to solve the problems. One of the main problems in KB-DSS is the acquisition and management of knowledge from the network. The knowledge base should contain the knowledge and the experience of the previous problems, as well as the tools to access this knowledge that has to be easy to create, maintain and update. For this reason, the use of Knowledge Management techniques is inevitable.

2.3. Knowledge Management

Management and control of networks is one of the main approaches of the companies during the past years. Large amount of information is stored in databases and this data should be analyzed, understood and processed in order to provide a success answer in troubleshooting issues. Systems need techniques able to manage information and provide expertise to the decision-makers. When a problem is identified, technician must have the knowledge information technologies engaged in the incident to solve the problem in the shortest time and cost for the company. During the lasts decades, the use of Knowledge Management (KM) techniques has been inevitable in systems like Help Desks, Fault diagnosis systems and Tickets systems.

KM is a technology that increases our understandings and helps the organizations to make decisions and solve problems more effectively by providing strategy, process and technology to spread information and experiences. Usually reasoning techniques are included in decision support systems to manage the knowledge. In this case it is organized according to a context specific to the system and how its agents perceive the problem domain. The most common reasoning techniques used in decision support systems are:

a) **Case-based reasoning (CBR):** Case-based systems use previous knowledge and experience to solve new problems by remembering past situations and reusing its solution and lesson learned from it. The knowledge base of a case based system represents situations or domain knowledge in the form of cases and the inference engine uses case based reasoning method to solve new problems or to handle new situations.

b) **Rule-based reasoning (RBR):** Ruled-based systems represent the domain knowledge with set of rules and suggest a solution or conclusion of a problem. In RBR the problem is solved based on a set of rules provided previously. A rule based system has one more component, which is known as working memory, in addition to knowledge base and inference engine. In these systems, the inference engine receives a problem from the working memory and provides the reasoning result to the working memory.

c) **Hybrid (a combination of CBR and RBR):** Developing system using one or more methods is the most advance in AI. The use of CBR and RBR methods in the same system allows a better simulation of human intelligence and a more efficient system. In the problem solving process, case based reasoning uses solutions that was solved in similar past problems whereas rule based reasoning solves problems from scratch though similar problems had been solved previously. At the end, a better solution will be provided.

Utilizing KM has been increasing within the past decades among organizations as a result of the promotion in creating, sharing and leveraging of the organization. With the increase of domain complexity, accelerating market volatility, intensified speed of responsiveness and diminishing individual experience, the integration of KM concepts and technologies is inevitable these days. For this thesis work, the Knowledge Management technology that will be used is Cased-based Reasoning (CBR).

2.4. Case-Based Reasoning

CBR is included within the fields of Artificial Intelligence (AI). During the last years, it has been used as a machine learning method. The origin of CBR comes from the work of Roger Schank and Robert P. Abelson in 1977 at Yale University. In this work, they proposed that our general knowledge about situations is recorded in the brain as *scripts* that allow us to set up expectations and perform inferences. The first system that used a Case-Based reasoner was called CYRUS, it was developed by Janet Kolodner in 1983 at Yale University (Schank´s group) and she based her work on Schank´s *dynamic memory model.* Based on Kolodner´s work, there are other CBR systems like MEDIATOR (Simpson, 1985), PERSUADER (Sycara, 1988), CHEF (Hammond, 1989), CASEY (Koton, 1989), or JULIA (Hinrichs, 1992). [1]

As an alternative, Bruce Porter and his team at the University of Texas in Austin designed the PROTOS system. This system used a heuristic classification and combined general domain knowledge and specific case knowledge into a single case-memory model.

During the lasts decades, CBR methodology has been included in a large number of application domains, for example, it can be found in Help-Desks[19], Decision Support Systems[14], Fault Management [15][16], Call centers [20], diagnosis and treatment systems in medicine [3][4] or web applications.

2.4.1. Definition

CBR is a cognitive method using in machine learning and in AI technologies. The principal aim of this method is to use past experience and knowledge of a problem domain in order to solve new problems. CBR compares the new problem (or case) with all the similar problems experienced in the past and adapt the solution for the new one. Instead of using all the knowledge of a specific domain problem, CBR is able to use only the information needed of previous experience. [2]

A new problem is solved remembering past similar situations and reusing the information in the new problem situation. As in the human solving problems case, where we are able to remember the past situations, in a CBR system is very important to store the new problem as soon as it is solved and make the information relating to it immediately available for a future problem. CBR is also a predictive system and it is quite important the learning process in this systems. It should be able to learn from experience and other methods.

This is a typical example of humans when solving problems: a woman wants to bake cake, but after one hour in the oven, it isn´t baked. What is the problem? She starts to think if the problem is the oven, or the ingredients… Suddenly, she remained another time that she had this problem because she forgot to use baking powder. Once she remembers her mistake, uses baking powder in the new case, and after 20 minutes, the new cake is baked and delicious!

In this example, the woman uses her memory to find a solution for the problem she has. She uses her previous knowledge baking cakes and adapts the past solution, in her new problem. She is using a case-based reasoning process to solve the problem.

In a CBR system, we can differentiate two important parts: the case, that means a situation or problem, and the case library, where are stored all the past situations or problems are stored.

2.4.2. Case

A case is a piece of knowledge representing an experience. It contains the details of the situation that we have experienced. At the end, the case has all the features to describe a problem.

The cases could be described in different ways, but all of the cases have the same parts: there is a *problem* description, which has a set of features, and there is a *solution*, which solves the problem. The case structure is often visualized in term of the *problem space* and the *solution space*.

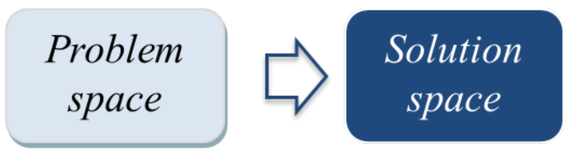

Figure 2-2: Case structure

When a new problem case appears, it is compares with other problem to look for a similar problem description. All the problems are in the problem space, so this new problem is also included in this space. To provide the solution for the problem, CBR system matches the problem part of the case and adapts a solution to the current situation.

2.4.3. Case Library

Once the new case is solved, it is automatically restored with all the past cases in order to have all the information to use in the future problems. In a CBR system, this knowledge is retained in a specific database called *Case Library.* All the problems and their solutions are stored in this case library.

The case library, as the human memory, has the information of the previous cases and the new ones, so it is quite important to define the descriptors of the library, for example, name, description, cause, solution, etc. of all the problems.

2.4.4. The Case-Based Reasoning Cycle

In 1994, Aamodt and Plaza introduced the CBR cycle, called “the four RE´s”, to describe how a CBR process solve a problem. In this cycle, they defined the following steps: *Retrieve, Reuse, Revise and Retain.* As they described this cycle: “A new problem is solved by retrieving one or more previously experienced cases, reusing the case in one way or another, revising the solution based on reusing a previous case, and retaining the new experience by incorporating it into the existing knowledge-base (case-base).” The CBR cycle is presented in Figure 2.4-2. [2]

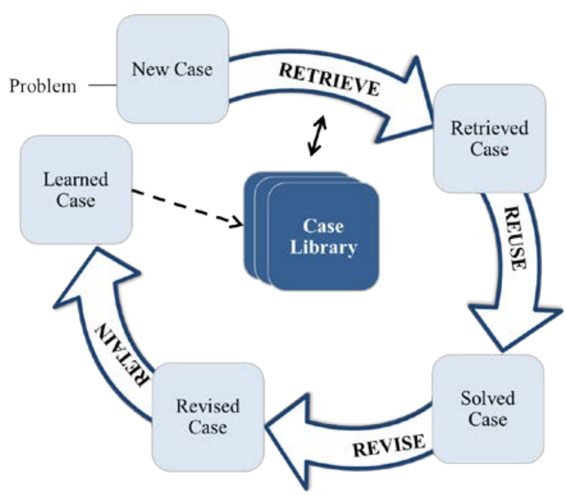

Figure 2-3: CBR cycle [2]

1. *Retrieve:* when a new problem (*new case)* appears, it is used to obtain a case from the *previous cases* that helps in our new situation. This step has an important application because it is the moment of recollect all the information relative with the case and may define the next step to solve our problem. The *retrieved cases* are compared with the new case matching the features and applying similarity matching algorithms. We could obtain several *retrieved cases* with a high probabilistic, but it is the moment of decide which is the best for our problem.

2. *Reuse:* once we decide the best previous case, the *retrieved case* is used in the reuse step. In this step, the *retrieved case* is combined with the new case to adapt both problems and find a suitable solution for the initial problem.

3. *Revise:* the propose solution is verified and tested for success. In this moment, we can use the solution in our decision support system and determine if it is correct to solve the initial problem.

4. *Retain:* if the solution has been succeed, we must store this experience for future use. The new case is retained as *confirmed case* and the system acquires a new learning process.

2.4.5. Knowledge

It is impossible to design a CBR system without knowledge. Developing CBR systems requires a systematic development of knowledge models by defining the requirements and building the models itself. CBR system´s knowledge can be divided in four different knowledge containers, as the model introduced by Ritcher: *vocabulary, similarity measures, adaptation rules* and *case base*. [30]

Vocabulary

The vocabulary contains the range of values used to express knowledge and defines the classes and attributes for query and case description. In a CBR system, the vocabulary covers the values that the system can deal with. Attributes can be numerical, symbolic, strings, textual or other. It also specify the language application, for example, this could be define the range of values (maximum, minimum) in numeric attributes or a list of values in symbolic attributes.

Similarity Measures

Similarity measures are the dependence between two values and describe how cases are related to each other. The assessment of the similarity between two cases is calculated by function depending of the attribute. It is possible to differentiate two types of similarity measures:

1. *Local Similarity Measures:* similarity method depends on the given data and available information. That means each attribute different similarity measure.

2. *Global Similarity Measures:* defines the relations between attributes. The weight of each attribute indicates its relevance within the case.

Adaptation rules

The Adaptation rules or Adaptation Knowledge gives information about how the cases can be used to fix the problem during the reusing step. This is often described by rules.

Case Base

As it was described before, a case is a piece of knowledge representing an experience. It instantiate attributes describing the problematic situation as well as a solution description. The collection of previous experienced cases are stored in the case base, also called case library.

2.4.6. When to use CBR in a system?

As it was discussed before, Case-Based Reasoning is included in many application domains, but it is important to define when CBR could be included during the development of a system. According to the application domain research, using CBR in systems can offer better results in the following cases:

- When monitoring and constant maintenance are required in a system. For example: systems using alarm detection and fault management networks.
- When previous knowledge helps for the new problem and previously successful solution exist. For example: systems using alarms.
- When similar problems show up often. For example: systems with no more of 7-8 possible problems and similar solutions.
- When support is needed to make a decision. For example: Decision Support Systems.
- When learning from new problems makes the system better. For example: maintenance systems.

- When many parameters interfere in the status of the system. For example: diagnosis systems.

2.4.7. Benefits of CBR

The benefits that Case-Based Reasoning provides in a system can be expounded as:

- CBR systems reduce the knowledge acquisition effort as it is possible to propose solution using the previous experience. It is not necessary to understand the problem domain because CBR allows the system to make assumptions and predictions based on what worked in the past without having a complete understanding. Making assumptions the system fills in incomplete or missing knowledge based on what his experience tells him, and goes on from there.

- CBR can propose solutions to problems quickly, reducing the time necessary during the evaluation of the system. Require less time and effort to solve a problem than an expert system.

- CBR gets more than one solution for a specific problem. When a CBR system proposes solutions, the user can select the best solution for the existing problem between several possible solutions.

- CBR uses exiting data and can learn from experience when a new case appears. Remembering previous experiences is particularly useful in warning of the potential for problems that have occurred in the past, alerting the engineer to take actions to avoid repeating past mistakes.

- Case-based reasoning gives a means of evaluating solutions when no algorithmic method is available for evaluation. Using cases to aid in evaluation is particularly helpful when there are many unknowns, making any other kind of evaluation impossible or hard. Instead, solutions are evaluated in the context of previous similar situations.

- Improve problem solving performance through their ability to use analogy. Cases helps the system to focus its reasoning on important parts of a problem by pointing out what features of a problem are the important ones.

- As a Learning Machine, the system can learn from the previous mistakes and update the case library using learning methods when a problem hasn´t been reported before. When no information is stored from a problem, the system use the previous data and evaluates the solutions from previous situations to proposed a solution and learn from the past.

- CBR is an easily understandable method and high acceptance for the user.

On the other hand, the most important limitation of CBR systems is the needed of previous knowledge to perform the system. Other possible limitation is the required of human interaction to use the CBR tool because the system offer a possible solution but it is the user who gives answers to the questions and the final solution.

2.5. Case-Based Reasoning Tools

When developing a CBR application, a large number of steps have to been account: first of all, collecting data and previous knowledge from the system and analyzing them to build a knowledge model and a suitable case representation; afterwards, designing an accurate similarity measure and implementing retrieval functionality; and finally, implementing user interfaces.

During this steps, a high effort is required during the knowledge acquisition, features extraction and, finally, case representation. The case representation should contain the less information needed to understand the problem and give a suitable solution using the CBR system. Analyzing data is one of the longest steps during the development of the application, but implementing the CBR application also requires a time consuming software engineering process and a lot of specific experience beyond pure programming skills.

Fortunately, in our days, engineers have available tools to build CBR system that support them during defining application analysis methodologies and provide assistance during the design and implementation of the system. In this paper, two of these tools will be presented: the first of these tools is the Open-Source CBR tool *myCBR*, developed at the German Research Center for Artificial Intelligence (DFKI); on the other hand, the platform *COLIBRI*, developed at the GAIA- Group of Artificial Intelligence Applications at Universidad Complutense de Madrid (Spain). [][]

In the following sections, a brief description of the previous mentioned tools will be provided to compare and analyze the better tools to use during the development of the CBR system proposed during this thesis work.

2.5.1. The myCBR Architecture

myCBR is an open-source similarity-based retrieval tool and software development kit (SDK). With *myCBR Workbench*, you can model and test highly sophisticated, knowledge-intensive similarity measures in a powerful GUI and easily integrates them into your own applications using the *myCBR SDK*.

Figure 2-4: myCBR [http://www.mycbr-project.net]

myCBR Workbench offers a GUI to rapidly prototype a CBR knowledge model of almost any given domain. The weight of the development tools available within the *myCBR* is aimed at the development of cases for structured CBR as well as on the ability to encode as much domain knowledge into the similarity measures of the knowledge model as possible. Starting from the knowledge gathered within a domain to be modelled the knowledge engineer (KE) defines the main concepts and their attributes to form up an initial case structure. In *myCBR* one it is possible to model concept hierarchies or 'part of' relationships. Available attribute types are integers, floating points, doubles, and symbols. After assigning the desired attribute types and value ranges the KE can model the similarity measures for attribute similarities. *myCBR* offers a range of graphical interfaces to assist the KE in modelling knowledge-intensive similarity measures. It also offers a tool to test the retrieval step of the CBR system.

The *myCBR 3 Software Development Kit (SDK)* gives a simple-to-use data model on which applications can easily be built. The retrieval process as well as the case loading, even from considerably large case bases, are fast and thus allow for seamless use in applications built on top of a *myCBR* knowledge model.

2.5.2. The *COLIBRI* Platform

COLIBRI is a platform for developing Case-Based Reasoning (CBR) software. Its main goal is to provide the infrastructure required to develop new CBR systems and their associated software components. *COLIBRI* is designed to offer a collaborative environment where users could share their efforts in implementing CBR applications. It is an open platform where users can contribute with different designs or components that will be reused by other users.

As a platform, *COLIBRI* offers a well-defined architecture for designing CBR systems. The platform has been designed in a two layer architecture: the *jCOLIBRI framework*, which is the is the white-box layer of the architecture for developing CBR applications in Java, and the new *COLIBRI Studio*, where the user can find several development tools during the implementation and sharing of new CBR systems and components.

jCOLIBRI is the he bottom layer of our platform and includes several features to define an interface between the data and the organization of the cases, retrieval methods as Nearest Neighbor, reuse and revision methods and maintenance. On the other hand, *COLIBRI* Studio is the implementation of the top layer of the *COLIBRI* platform. It provides the visual builder tools required to generate CBR systems without dealing directly with the source code. It is built on top of the *jCOLIBRI* framework and enables the composition of its CBR components.

Figure 2-5: jCOLIBRI platform [http://gaia.fdi.ucm.es/research/colibri]

COLIBRI supports the development of several kinds of CBR systems: as standard CBR systems, textual CBR, knowledge-intensive, data-intensive or distributed CBR applications [11]. It also includes evaluation, maintenance and case-base visualization tools. Many of the components available have been developed by third-party research groups and contributed to the platform to be shared with the community.

3. STATE OF THE ART

This chapter explains systems using Case-Based Reasoning to provide decision support in solving a problem. CBR methodology has been applied in a large number of application domains, for example, it can be found in Help-Desk[19], Fault Management [15][16], Call centers [20], diagnosis and treatment systems in medicine [3][4] or web applications.

3.1. Case-Based Reasoning in DSS

Several variations of case-based reasoning process model exist in the literature (Riesbeck and Schank, 1989; Hammond, 1989; Kolodner, 1993; Aamodt and Plaza, 1994), but the most practical one is that introduced by Aamodt and Plaza with 4-R cycle: retrieval, reuse, revise and retain. [2]

Knowledge-based systems (KB-DSS) are used primarily for fault diagnosis and maintenance in networks as presented in papers [15] and [16], but earlier research are focus on Customer Services and Help-Desks systems as projects of papers [19], [14], [20] and [21]. At the same time, CBR researchers have realized the limitations of traditional CBR systems and have begun to move towards hybrid CBR architectures, incorporating various techniques and concepts to enhance the CBR process and performance. Example of integration of CBR and other reasoning methods could be read in papers [21] and [23].

In the following sections, related work using CBR techniques will be discussed. First of all, two examples of CBR techniques focus on fault and alarm management in computer networks will be described. Following, a description of CBR integrated in Help Desks and Services Desks will be explained.

3.1.1. CBR in Fault Management

One of the first uses of CBR techniques were in system specialized in control and diagnosis of networks failures. Nowadays, the complexity involved in computer network fault management demands a great amount of information about the involved technologies and their associated problems. Networks often consist of thousands of interconnected nodes and the possibility of a failure is quite high. The solution for the companies to detect and control the problems that could occur in the network is using support system as Trouble-Tickets systems [15] or Alarms systems [16]. When a large number of information has to be managed by the system and the potential problems need a successful solution, a system able to reuse the information stored is required.

Many companies have been using Trouble-Tickets systems to store the occurred problems, actuating as an historical memory of the network. However, the increased of the number of problems and the complexity required in the diagnosis of the current networks have required the development of new systems able to use all the information about the previous problems during the diagnosis of a new problem. As in paper [15] presented Trouble-Tickets systems aid the managers in the task of monitoring problems occurring in a network storing the historical memory of the faults of network. Using

CBR techniques in fault management networks, new experts systems take in account the knowledge accumulated in the Trouble Ticket Systems to help the managers in the diagnosis of a new similar situation and propose solutions for the proposed problem. In this project, the system called DUMBO is the solution for the proposal problem.

DUMBO System [15] uses CBR system that has been developed upon the architecture and technology of traditional Trouble-Ticket systems. In this example, the system uses the structure of the previous ticket system but including new functionalities in order to add some reasoning process for the retrieval of cases. The cases, represented as tickets, have two parts: the “problem description” and the “solution description”, and also some additional features are added to compare with other tickets during the similarity evaluation.

Furthermore, in paper [16] a CBR system is integrated within an alarm correlation system in a telecommunication networks. In the network management domain, alarm correlation is often used to help in the real-time diagnosis of faults and fault localization. Reported alarms from a single network are correlated to reduce the amount the information reaching the operators, but a system able to search for similar alarms was required. The proposed system is able to filter and correlate alarm using CBR methods and offer solutions for the networks at higher speed than the previous one. Controlling faults with alarms and CBR system increase the precision of the detection of errors and solving failures.

3.1.2. CBR in Help-Desks

Help-Desks are installed in many companies to help customers, which have problems with the company’s products and services. But customers are not the only users of Help-Desks, field engineers and managers usually use them to search similar problems which could help to solve the new problems. In short, solving a problem by retrieving a similar problem and applying its solution to the current problem is the theory of Case-Base Reasoning (CBR). Therefore, using is a suitable technology to assists Help-Desks in their daily routine. CBR is increasingly adopted for the development of intelligent systems to manage various types of Help-Desk support operations. Latest researches have been focused on use Knowledge Management structures in Help-Desks systems by integrating Case-Based Reasoning and Service Desk structures as we can see in the projects presented in papers [13], [14], [17], [18], [19], [20] and [21].

By adding Knowledge Management techniques, new systems are able to improve Help-Desks operation and ensure better and faster solutions for problems. The improvements of the system offer to the companies a more quality and effective customer service and allow a cost saving in solving problems.

One example of system that integrates CBR techniques and Help-Desk structures is the HOMER system [14]. As the author explains, The HOMER system was one of the application results of the INRECA-II2 project with its main goal to improve the development process of industrial Case-Based Reasoning applications. In this project, the system developed introduced CBR theory in the CAD/CAM Help-Desk used at DaimlerChrysler in Sindelfingen. The HOMER Help-Desk support system was able to manage the knowledge and offer several improvements for the previous Help-Desk system which operated in a level organized system: the first level for calls, second level as trouble-ticket system and third level for new manufactured products. Since the number of calls handled by help-desk personnel was really high and help-desk operation was a central and very time-critical operation, it was obvious that a new system was need. On the other hand, help-desk personnel didn´t give support in diagnosing the problem and could not serve as a knowledge repository. Thus, the Homer system was developed. In papers [13] and [14], a description about the HOMER system and an explanation of the features of the system and how was the process of development and integration in the previous Help-Desk is presented.

First step in the development of a CBR system is to decide how to handle the knowledge and define the method by which the problem-solving experience will be represented in the system. The HOMER system presents an easy structure for Knowledge Management that could be useful in our project. In this case, they decided to use an object-oriented approach to model the domain and discomposed the tickets according to a “failure”, “symptoms” and “solution” schema. Additionally, the HOMER system recorded administrative information about the author of the case and the time it usually took to solve the problem. As the author explains [14], the detailed decomposition of the cases was an important source for refactoring efforts during solving problems with respect to the domain model and, in case of recurring problems, it helped to find the root causes and save time. Using HOMER system, Help-Desk operators are able to access, re-use and extend the knowledge in a natural and straightforward manner.

On the other hand, paper [18] presents the integration of CBR techniques in a Help-Desk service operated manually by a small team of technical and help-desk personnel in a small bank. The system was called CBR Express. This project was started more than 10 years ago, but it may be helpful to understand the application of CBR in Help-Desks. In the small, operators had to handle with a large number of requests of service and information and also for problems. For faults, if the response is not automatically, the troubleshooting had to continue until it was solved. First of all, the team usually handled with urgent and mission-critical problems to the day-to-day operations of the bank and the steps in network fault diagnosis was to observe the symptoms, hypotheses derivation, run some checks or tests, and form some conclusions. The process was slow and hard. In this case, the help desk operators needed to diagnose the problem so that they can dispatch appropriate help. By using CBR, they tried to update the system to offer shorter responses and facilitate the process of knowledge acquisition and update.

During the system development, the first step was to design the case representation and the case library. For this example, the model of representation was based on partitioned a combination of discrimination and shared-featured memory organization network structures as Kolodner used in 1993. Cases were represented using the following parts: “case title”, “problem description”, “cases attributes” and “action”. It also used rules to associate cases. Integration the CBR system in the old Help-desk, the expert system helped to streamline the management of hotline operations and enabled the team to function at a high state of operational readiness in response to mission-critical problems, as the author concluded.

4. PROBLEM DESCRIPTION

This chapter describes the current system in the DSO and limitations of the current systems. Finally, the proposed solution of the problems in this DSO system is explained in the following sections.

4.1. Problem description

The electricity system connects power plants, network operators and electricity consumers and covers the processes and management of the production, transmission, distribution and storage of electricity. The electricity is produced by a transmission system operator and the DSO is the responsible of transferring electricity from the producer to the customer and distributing electricity from the transmission system to places of consumption.

In our days, the DSO maintains and repairs more than 60000 kilometers of power lines and 20000 substations which are controlled by the central system. In order to improve the quality of this system, the main objective for the following year is to use a system based on Smart Meters.

Smart Metering is a very promising technology that can substantially empower electricity customers to become active managers of their consumption. Smart Meters will improve the customer's knowledge about his/her electricity consumption thereby increasing customer awareness of energy end-use. Besides, Smart Meters will allow an optimization of the customer processes, making them more efficient and more reliable thereby leading to enhanced supplier switching and higher customer satisfaction. Lastly, Smart Meters will lead to an optimization of the overall electricity distribution infrastructure.

Using the Smart Metering technology, the DSO wants to automate as much as possible the information gathering and fault analysis process for the clean-up technician using analytics and knowledge reasoning. In a telecommunication domain, these kinds of systems are called network operation centers (NOC) and are responsible of network monitoring, incident response, communications management and reporting problems.

In a NOC, the different elements of the network report anomalies in the system generating alarms that are managed in the operation center. These alarms are stored in the system looking for the best solution for the existing problem. After an alarm, a NOC engineer will search for the cause of the problem in order to solve it, but sometimes it takes a lot of time and money. Looking for similar alarms and previous knowledge will help engineer to solve problems.

This thesis work proposed a decision support system based on CBR technology that uses previous knowledge to create the meter management tool that will support engineers during the customer's calls.

4.1.1. Current system in the DSO

The current operation center used by the DSO has meters in every end customer in order to record all the information about the current situation of the electricity network. The nodes of the network collect all the data about the meters from different sources and analyze the parameters sent by the meter defining the quality of the service. When a problem is reported, the field technician has to check all the data available in the system to solve this problem as soon as possible.

The existing system can be divided in three parts: the end customer, the call center and the field technician. The call center can communicate with the end customer and with the field technician by phone during the problem is being solved and the end customer can be in contact with the field technician either by phone or physical, as it is illustrated in Figure 4-1.

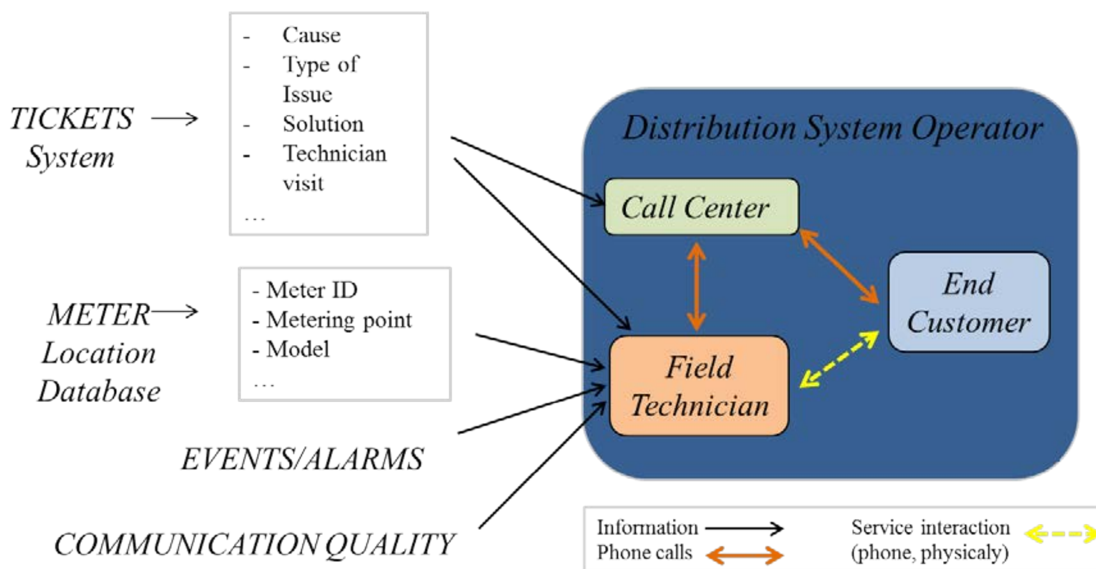

Figure 4-1: Current system

When users detect an issue in the system, they have to contact the call center and inform about the problem. First of all, the call center tries to solve this problem following a troubleshooting guide, the available data in the system and the help of the end customer. If the problem can’t be solved in this way, the call center contact with a field technician to solve the problem. The field technician uses the information gave by the user and also can use the data stored in the system. The data is constantly updated with the information of the meters.

There are different types of data in the network that the field technician can use to solve the problem. The available data is described as follow:

- *Tickets*: when a problem is solved, a ticket is created by the field technician. This ticket has all the information regarding the problem: customer information, meter information, cause of the problem, type of issue, solution or information about the solution procedure (e.g. if the technician went to the site or the problem was solved remotely). All the previous tickets are stored in the Ticket system.
- *Meter Information:* the information about the kind of meter it is also available: meter ID, location, model or kind of meter. In the Meter Location Database, it is possible to find all the information regarding the meters.
- *Events and Alarms:* the meters also send information about the status of the meter (power off, meter error, critical alarm…). In this case, all the events and alarms that occur in the meter are stored in one of the nodes of the network. The technician can use this information to search for previous problems reported in the meter.
- *Communication Quality:* The quality of the meter is analyzed using the communication quality of the meter. When this parameter is decreasing, the

technician can determinate that the problem is during the communication between the meter and the network and it helps to diagnose the problem.

The main problem of this current system is that there isn´t any correlation between the data and many different points should be checked to understand the problem in the meter. When a problem is reported, the technicians need a long period of time to solve usual problems and it is slow and expensive. For this reason, the objective of this thesis work is looking for a system which automates as much as possible the available information and helps the field technician during the fault analysis process to save time and money.

As it is mentioned before, the main drawback is that problem solving in the distribution network are sometimes slow and expensive, but there are other kind of problems that have been discovered in the system. First of all, using the existing system, the technician could be wrong and it means that a problem can't be solved properly; human errors also mean gaps in the tickets, missing data and incoherent information. Secondly, using different source of information, specific knowledge domain is requited to understand the data and correlate with the other sources. On the other hand, the old system have problem with meter located in specific areas as it means complicated fault location and difficult signal analysis.

4.2. Proposed solution

The new system should solve all the previous problems and store all the relevant information in the same node in order to facilitate the fault analysis and diagnosis when a problem is reported. This system has to help call center to solve the usual problem proposing possible solutions and using the previous knowledge. To define a system that uses knowledge reasoning, data has to be analyzed and defined in order to create a common structure for the whole system. Using information from all the nodes, it is possible to develop a system which could help call center and field technician with this issues.

The proposed solution for the problem described in the previous section is to include a DSS based on Case-Based Reasoning techniques that correlates all the information and propose possible solutions when a new problem is reported.

A system based on CBR techniques uses previous knowledge and information to propone solutions to the new problems, as it was explained in section 2.4. For the existing problem, the proposed solution is to use all the information from the current system to develop a DSS based on CBR where either call center and field technician could check previous and similar problems that helps in the new one. A representation of the proposed system is shown in Figure 4-2.

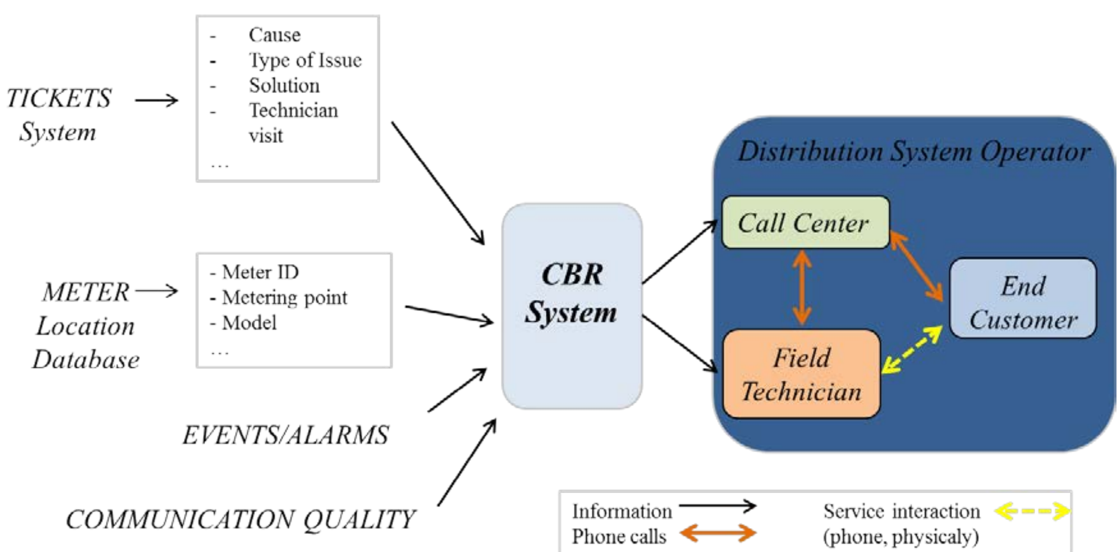

Figure 4-2: Proposed CBR System

Using this system, all the previous information stored in the system is reused to the new system. First of all, information from similar tickets stored in the Tickets system could be restored and information of the status of the meters could be used to analyze the failure.

The benefits of the proposed CBR system are several. The system will be develop using the existing data and it will be update using the future problem; a system based on CBR is to retain the new experience acquired from the meters faults. New source of knowledge will be useful to solve the problems in the network because CBR is also a machine learning method. Secondly, call center will give faster solution to the end customer. Call center will be able to solve problem without the required knowledge domain as a CBR can propose a solution using certain parameters. On the other hand, field technician will go on-site only in certain occasions, as the system will help during the diagnosis system or asking the user to do some tasks. It’s involved a cheaper solution to the company as the technician will travel only if it is required.

5. DATA ANALYSIS AND PRE-PROCESSING

This chapter describes the analysis and pre-processing of the data given by the company. This is one of the major and time-consuming parts of developing the proposed DSS. The objective of this chapter is to offer a detailed explanation of the existing data in order to understand the data and define the problems and solution reported in the electricity network. This analysis started cleaning-up the available data and defining the database of the use case. Once the relevant data was defined, it was time to analyze the Tickets system and figure out the problems in the electricity network and the solutions given by the field technicians to these problems. At the end, other files were also studied to search more features which help to detect problems in the network.

5.1. Initial analysis

After the problem description, it is time to start with the data analysis. The analytics started off having a first look of the files given by the company and looking for relevant information about the problems and the solutions reported in the Tickets system. More than 10K tickets and 130K events/alarms are stored in the available files, but a large amount of gaps and missing information are also found.

At the beginning, Tickets were analyzed looking for the main characteristics of the problems and its solutions. A ticket contains all the information related to the problem in the network and it is needed to define the features of the decision support system. The main features in the tickets are the meter identifier, the cause of the problem, the type of problem, the solution of the problem, and if the technician went to the meter or the problem was solved remotely.

On the other hand, using the available data from the events and alarms and the communication quality, information about the meter, warnings in the electricity network and the status of the communication between the meter and the monitoring network can also be figured out. More than 130K events and alarms were available, but the relevant events/alarms to analyze are those concerned with the problems in the Tickets file. The relation between the Tickets system and the events/alarms files is the meter identifier, also called EIC code. With this code, it is possible to find all the information about the meter as is the connection between all the files in the existing system.

Due to the large amount of data and the missing information in the files, it was impossible to define a relation between the problems and the solutions and a first selection of data was needed. In order to get a clearer database, the first step of the data analysis process was cleaning-up the data and selecting the relevant data. In the following section, it is possible to understand how this clean-up was carried out.

5.2. Cleaning up the data

During the analysis, the first step was cleaning the data. Once the table was smaller, problems analysis was performed. The condition to reduce the number of cases of study was to use only those cases were cause, solution or technician visit were reported in the tickets. The cases without these fields filled in are supposed “human errors” and discarded in the analysis of the data. The first cleanup gave a selection of 1198 to analyze.

Continuing with the analysis and looking for feature in the files, a second selection has been carried out. The number of tickets was very high and the missing information was still relevant. To define the relation between the problems and the solutions and understand how the company solved every problem, the second tickets selection was those tickets where the information in causes and solutions were filled in. In this case, the number of cases found was 457.

Using 457 cases, it was possible to start with a deep analysis of the data. During the data analysis, firstly the data stored in the tickets system was analyzed in order to describe the different problems and solutions that the electricity network can report. As it was mentioned before, the Tickets system stored the information about the different problems reported in the meters as cause, date, customer, technician name and solution can be found in this file.

The clean-up process was one of the larger steps because all the files were analyzed looking for the relevant features from the files and discarding that information which didn´t help to figure out problems or solutions. An important stage of this process was to describe each feature, define the type of the data and the values and decide if the feature was useful or not. In the Table 5-1, an example of the selection of the data is given.

Table 5-1: Feature description

Name	Description	Type	Value/Format	Useful
Key	*Ticket identifier*	*Text*	*KP- XXXXX*	***YES***
Summary	*Description of the ticket*	*Text*	*Free text*	***NO***
Issue Type	*Internal referral to*	*Text*	*Free text*	***NO***
Status	*Status of the ticket*	*Text*	*Open/Closed*	***YES***
Priority	*Priority of the ticket*	*Text*	*Minor/Major*	***NO***
Resolution	*Status of the problem*	*Text*	*Unresolved/resolved*	***NO***

All the fields found without relevant information or with gaps have been removed from the file. An analysis about the data has been also performed in order to understand every field of study. The cleaning-up helped to remove the irrelevant information or the data that the company doesn´t want to use in the system.

5.3. Ticket analysis

The analysis of the data stored in the ticket system helps to understand the kind of problems reported in the electricity network and the solutions that the technician gave to each problem. During the analysis, it was possible to define the relation between the problems and the solution in order to define the problem-solution schema.

At the beginning, all the tickets were analyzed to figure out the causes of the problems. The causes define the possible problems that could appear concerning the day-to-day operation of the electricity network in the DSO. The problems found during the ticket analysis can be caused of meter failures, networks problems, noise or other problems. In the 457 cases studies, the following causes were reported in the tickets:

1. *Connection restored:* During a while the connection wasn´t working but now it is restored. The problem is not identified because it was fixed itself.

2. *Customer due:* It is not possible to access the meter because of the customer.

3. *Main switch OFF:* The meter hasn´t communicated the last 3-4 days because the main fuse is turn off.

4. *Meter display failure:* The meter display is not working and it not possible to see the meter.

5. *Meter exchanged by the company:* The company owner of the meter has decide to change a meter and there isn´t any communication about it.

6. *Meter failure – Critical meter alarm:* The meter is reporting a critical alarm due to a problem in the meter.

7. *Meter failure – Other reasons:* Other kind of root causes. There is a problem in the meter but it not possible to identify the problem because it is not a display failure neither a critical alarm.

8. *Mobile network coverage:* There are issues with the signal, for example, GSM/GPRS issues. The network should be 3G, if the network is 2G it is needed to use an external antenna.

9. *Noise:* There is heavy interference between the meter and center of monitoring. The meter has been removed from the SLA/monitoring or is not working because of the noise.

10. *Other reasons:* Other kind of problems reported in the network but not differentiate in the ticket system.

Once the different causes were explained, it was possible to define the kind of possible problems in the network. The causes have been integrated in 5 different groups depending of the kind of problems. The distribution of the causes in this case can be understood in Table 5-2.

Table 5-2: Causes definition

CAUSE	Description	#
Main switch	The main switch is OFF	**170**
Meter failure	Problems reported in the meter	**109**
Other reasons	Different problems	**99**
Mobile network coverage	Problems in the network	**65**
Noise	Heavy interferences in the meter	**14**

In Figure 5-1 it is possible to see the distribution of each cause in the system. The main important problems are the main fuse problem when it is OFF and the meter failures. In the other reasons groups are located all the problems reported only in few occasions as customer due problems, slots of time without connection and actions that the company decides to perform in order to improve the quality of the system , for example, restart or change meters.

After defining the causes in the network, the other important part of the data analysis was the definition of the possible solutions used to solve the problems. The solutions are defined according to the decision of the technician. In some cases, the customer can solve the problem immediately following the instruction of a technician or from the customer center, but in other cases the technician visit is required.

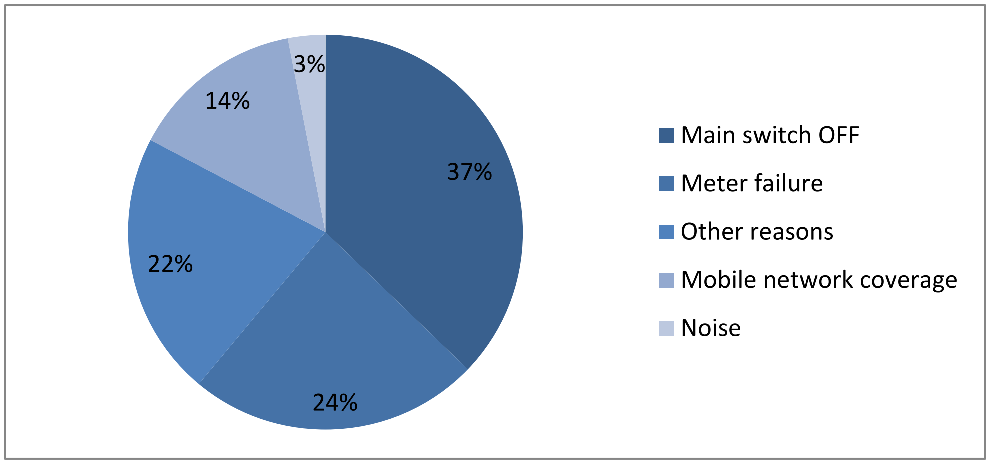

Figure 5-1: Causes distribution

When a problem is reported, the customer should call to the customer center in order to contact with a technician and fix this problem. According to the solution found in the tickets system, the following 5 solutions can solve the previous problems:

1. *Antenna installation:* The technician installs a new antenna when a problem with the mobile network coverage is reported.

2. *Main Switch ON:* If the main switch is OFF, the solution is to turn the main switch on. Turn off and turn on the meter again to restart the communication. The meter restarts automatically when the meter turn on again.

3. *Meter restart:* If there is a problem with the meter, the first step is to restart the meter. This solution should be enough if the problem doesn´t affect the network. If the problem is not solved after this solution, a technician opinion is required.

4. *New meter installed:* If there is a non-curable problem or the meter is broke, a new meter is necessary. In this case, the technician installs a new meter.

5. *Remove source of noise*: If the meter can´t be monitored because there are heavy interferences in the network, the unique is solution is to remove the noise. In this case, the technician has to find the source of noise and remove it.

As it has been described, depending of the problem, the solution can be one or other. In order to understand the possible solutions proposed for each problem, a study about the percentage of use each solution according the different problem has been also done.

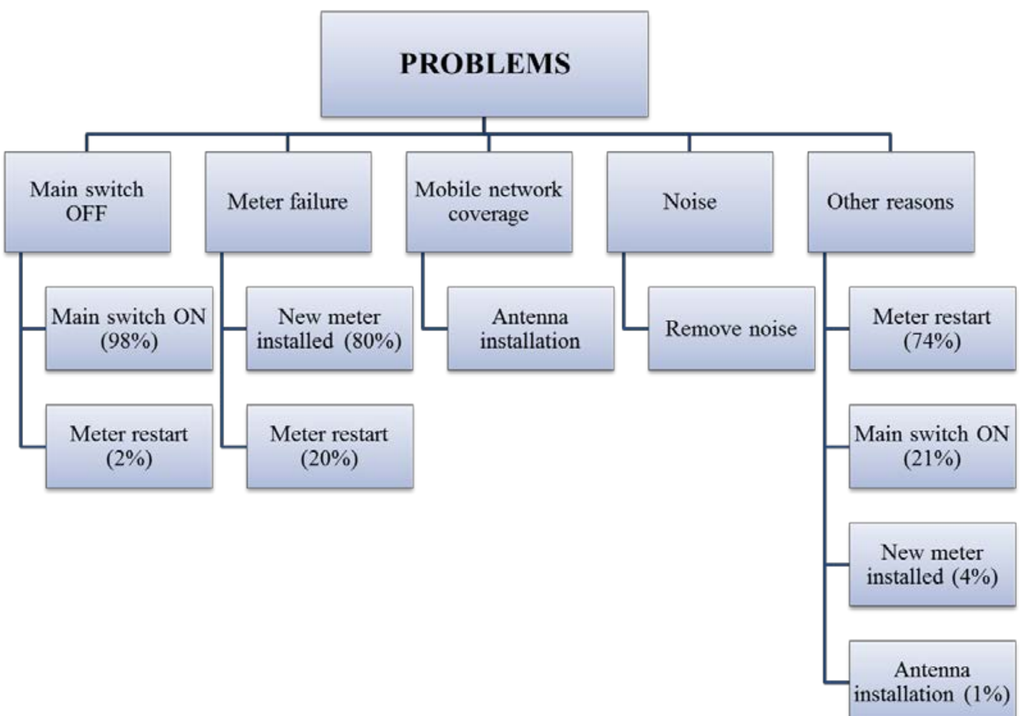

Figure 5-2: Relation Causes - Solutions

In Figure 5-2 it is possible to observe the relation between the causes and the solutions.

According to the previous figure, the solution when a problem in the mobile network is reported is to install a new antenna and the problem should be fixed. In the case of noise, the solution is also easy, the technician needs to find the source of noise and remove it to recover the meter communication. On the other hand, when another kind of problem is reported, the solution may depend on other parameters. In this cases, the way of solve a problem should be define. Now it is time to study how the customer center proceeds when a call reporting a problem is received.

In order to understand how a problem can be solved, the following workflow has been created. In Figure 5-3, it is possible to understand the possible problem depending on the location of the problem and the proposed solution in each case.

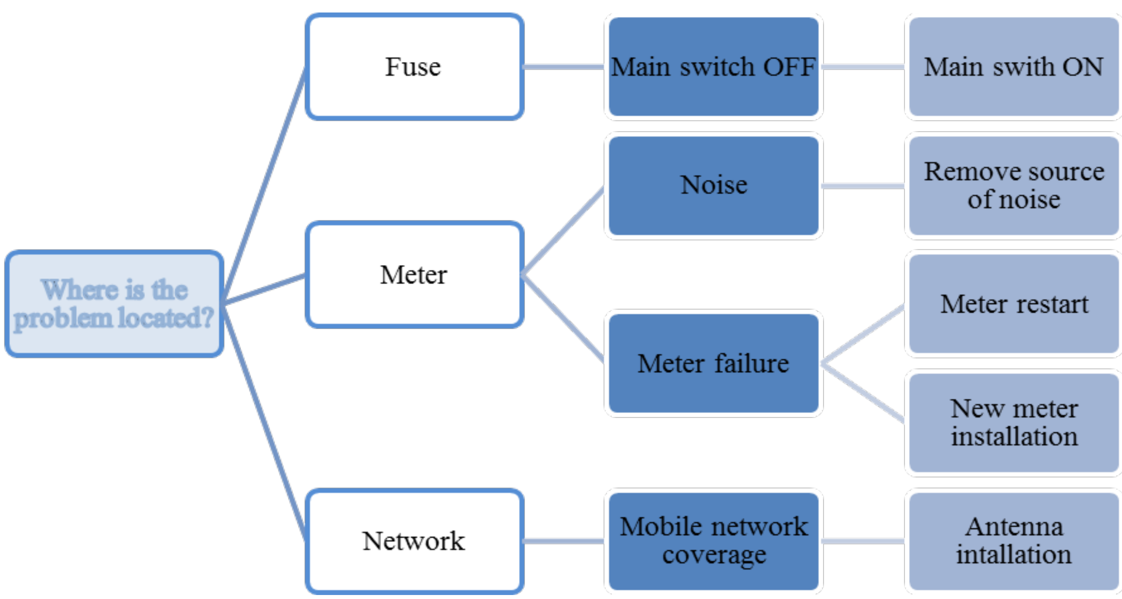

Figure 5-3: Workflow in the electricity network

5.4. Events and alarms analysis

In the file where the events and alarms are reported, the most relevant information is to clarify if there were any warning before the problem was reported. Extracting the events and alarms the 30 days before of the problem, it is essential to analyze if the meter had problems that weren´t reported during the previous problem and understand the kind of problem. In Figure 5-4, some examples of possible events and alarms in the meter are presented.

id	type_code	reason
32791589	3.26.0.85	Quality log: Power down
32791590	3.6.29.16	Alarms and events: Terminal cover closed
32791630	3.6.29.212	Alarms and events: Terminal cover removed
32791648	3.26.0.216	Quality log: Power up after short power cut

Figure 5-4: Events and alarms example

5.5. Communication quality analysis

During the communication quality analysis, the main idea was to define if it was useful to include in the proposed system. The information stored in the data warehouse node gives the history of the communication quality of a specific meter. As the function of the communication quality is determinate if the meter is receiving all the messages that have been sent, a filter is need to define if there is a problem in it. In this case, the filter to take the communication quality is 30 days before of reporting the problem.

From the communication quality file, it is possible to identify if the meter had previous problems when a problem was reported. In the following example, it is possible to see the average of the communication quality and when the meter stopped to receive communication. To calculate this parameter, the system need to know the number of messages sent and the number of messages received. The formula is:

$$Comm.Quality = Message\ Sent\ /\ Message\ Receive$$

Below the evolution of the communication quality before the problem is presented:

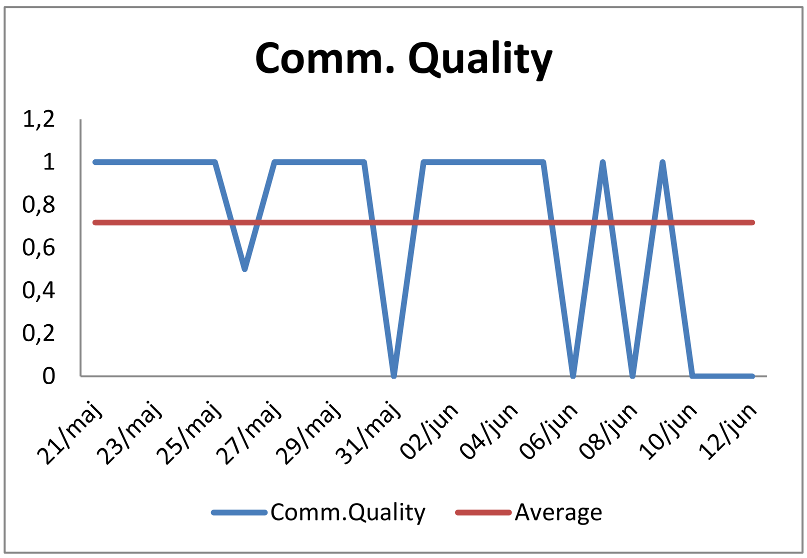

Figure 5-5: Comm. Quality Example

6. CBR SYSTEM DESIGN

To develop a system based on CBR techniques it is required to complete all the steps of the 4R´s cycle described by Agnar and Plaza: Retrieval, Reuse, Revive and Retain. However, implementation of the CBR cycle has not been included in this work; only an initial retrieval approach has been discussed using a tool *myCBR*. This chapter includes a detailed description of the stages followed to define the final case library for the CBR system and also an overview of the functionalities of the tool *myCBR* used to design the initial retrieval approach of the proposed CBR system.

6.1. Stages in the CBR System design

In order to clarify the functions needed in developing a DSS in this domain, the thesis work started with study of the current electricity network operation center and the analysis and definition of the problems that the company is using now-a-days. After understanding the problem domain and analyzing the data given by the company, it was to time to continue with the extraction of the important features in order to define the cases in a case library for the proposed CBR system.

As it was mentioned in the CBR description, it is impossible to design a CBR system without knowledge. For these reason, one of the most important step during the development and design of a CBR system is the knowledge acquisition in order to build the knowledge containers introduced by Ritcher: the vocabulary, the similarity measures, the adaptation rules and the case base. [30]

In this thesis work the main objective is to build an appropriate case base and the vocabulary of the system using the relevant information from the electricity network. On the other hand, a description about the similarity measures is also required to evaluate the system. However, adaptation rules have not been included in the development of the proposed CBR system.

Once the data was analyzed and the distribution of the problems was defined, it was time to start with the design and develop of the case library for the proposed CBR system. This case library should contain the information regarding the problem and solution domain as well as other features which help to propone a solution using the tool *myCBR* when a new case is defined. During the development of the case library, it was discussed the extraction of the important features and the case representation. After defining the case, a case library was built using those cases where both causes and solutions were filled in the Tickets system. The stages followed the development of the small prototype of the proposed system are shown below.

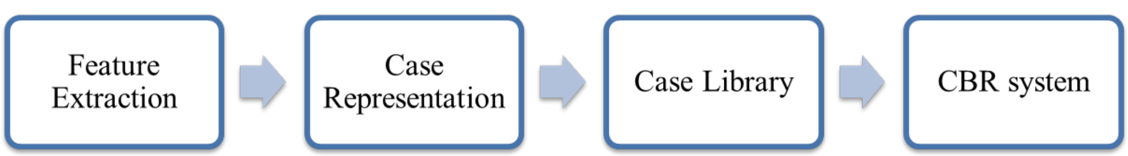

Figure 6-1: CBR classification

1. *Extraction of the features:* Extracting useful features to represent cases is one of the important steps of a CBR system. Once the analysis data was completed, the important features from the tickets were selected. The selection of the useful features to search similarities between the cases was one of the largest steps during this project.

2. *Case Representation:* The main goal of the feature extraction was to define the case representation of our CBR system. To choose the best representation of the problem-solution domain, similar systems using CBR methods were studied. Finally, it was decided to follow a structure which involved *problem description, symptoms* and *solution description.*

3. *Building of the Case Library:* Once the case representation was defined, it was time to build the case library for the proposed CBR system. The case library is the knowledge container of the CBR system and stores all the relevant information about the problem-solution domain. The database of cases was built manually using *Microsoft Excel 2010* and a *CSV* (*Comma Separated Value*) format.

4. *Developing of CBR system:* The proposed CBR system doesn´t include the complete CBR cycle, it was only developed a prototype of CBR system using the case library in the tool *myCBR* to define the retrieval approach of the system. *myCBR* includes functionalities to import a case library from CSV files and to define the similarity measures between the cases. One of the most important steps in these stages was the definition of the similarity measures: first, similarity measures for each attribute were chose (*Local Similarity Measures);* and secondly, the weights of the different attributes and the *Global Similarity Measures* were defined.

The following sections detail the stages of the project and describe the development of the prototype CBR system proposed to the company at the ending of this thesis work.

6.2. Feature Extraction

Once the data analysis was completed, the next step was to extract the features from the files to define the case representation of the system. During the data analysis, a first clean-up of the data were done when all the irrelevant information and missing data were removed from the initial files. After filtering the data, the data was much clearer and it was possible to decide which features could be useful for the final case representation.

In order to build a case library for the system, it was essential to choose the useful field from the available files and extract all the data to create a unique file. To extract the important features and define the case representation, a link between all the files is needed. In the available data, the link is the EIC code which connects the information from the tickets system, the meter database location and the data warehouse. Using the EIC code the user can obtain all the information gathering the tickets, meters, events and alarms and other parameters as the communication quality. However, for the thesis work the EIC code was used during the extraction of the important features from the tickets system and the meter information.

A schema of the data used for extracting the features is shown below in Figure 6-2.

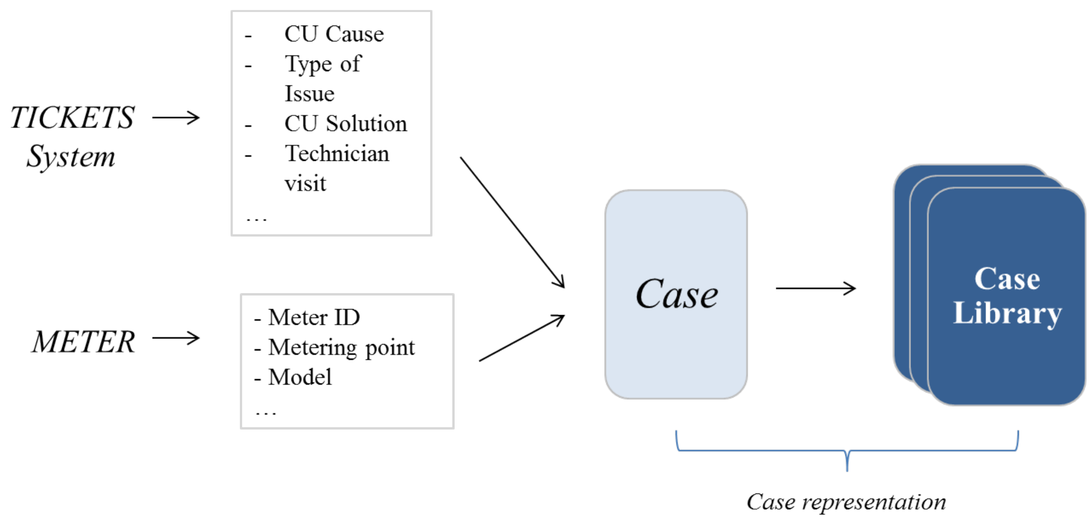

Figure 6-2: Feature extraction schema

As it was mentioned before in the previous chapter, during the cleaning-up of the available data a description of each feature was carried out in order to decide if the feature could be useful or not in the case representation. Following this classification, the following were extracted from the tickets system and from the meters.

The feature extraction process was done manually using Microsoft Office 2010 and using the EIC code to collect the data from the different files. In the future, a feature extraction algorithm will be developed to include in the system.

6.2.1. Extracted features from the tickets system

The existing tickets system has all the information of the problem and also the information of the solution, as it was described in previous section. In order to get a basic description of the problem domain to use during the retrieval step in the CBR system, the extracted features from this file are presented in Table 6-1.

Table 6-1: Extracted features from the ticket system

Name	Description	Type	Value/Format
Key	*Ticket identifier*	*Code*	*KP- XXXXX*
Status	*Status of the ticket*	*Symbolic*	*Open/Closed*
Created	*Date of creation of the ticket*	*Date/hour*	*YYYY-MM-DD HH:MM:00*
Updated	*Date of update of the ticket*	*Date/hour*	*YYYY-MM-DD HH:MM:00*
Linked Issues	*Other tickets relating the problem*	*Code*	*KP-XXXXX*
Type Hits	*Location of the problem*	*Text*	*DC–Data Concentrator/ PLC –Power Line Communication/ Fault Doubt*
Cause	*Cause of the problem*	*Text*	*Causes distribution**
Deadline	*Date of the deadline of the ticket*	*Date/hour*	*YYYY-MM-DD HH:MM:00*
Priority Date	*Problem priority during the date*	*Symbolic*	*High/Medium/Low*
EIC code	*Meter identifier in the network*	*Code*	*38ZEE-NNNNNNNN-X*
Type of customer contacts	*Information provided by the customer during the call reporting the problem*	*Text*	*Meter lectured/ Meter malfunction suspected/ DC/ info/Main fuse/ PLC*
Issue Type	*Type of problem*	*Symbolic*	*From CC/Ericsson's internal referral to*
Satellite ID	*Identifier of the Satellite connected with the meter*	*Symbolic*	*PRE-SAT/SAT/SLA*
Technician´s visit	*If the technician went to the meter station or not*	*Symbolic*	*Yes/No*
Solution	*Solution of the problem*	*Text*	*Solutions distribution***

**<u>Causes distribution:</u> as it was described in the previous chapter, during the data analysis 5 different possible causes were defined: Main Switch OFF, Meter failure, Mobile network coverage, Noise and other reasons.*

***<u>Solutions distribution</u> as it was analyzed in the previous chapter, 5 possible solutions were described for the causes: Main Switch ON, Meter restart, New meter installed, Antenna installation and Remove source of noise.*

6.2.2. Extracted features from the meter

The information about the meter was extracted from the file where the events and alarms were reported. To define the parameters of the meter, the features presented in Table 6-2 were selected.

Table 6-2: Extracted features from the meter

Name	Description	Type	Value/Format
Meter ID	*Meter identifier*	*Numerical*	*XXXXX*
Metering point ID	*Location of the meter*	*Numerical*	*YYYYY*
Serial number	*Serial number of the meter*	*Numerical*	*ZZZZZZZZZ*
Meter created at	*Date of creation of the meter*	*Date/hour*	*YYYY-MM-DD HH:MM:00*
Meter updated at	*Date of update of the meter*	*Date/hour*	*YYYY-MM-DD HH:MM:00*
Model	*Model of the meter*	*Symbolic*	**3 different models*
Remotely readable	*If it is possible to read the meter remotely of it is needed to go on-site.*	*Boolean*	*0/1*

6.2.3. Feature extraction representation

An example of the extracted features to build the case library is presented below. A representation of the information from the tickets, the alarms/events and also the communication quality of the example can see in the example.

Features extraction example

In Figure 6-3, there is a representation of the features extracted from the different files. In this example, the EIC code=”38ZEE-00003958-8” was used and it is possible to analyze the information from the files.

TICKET 1	
Key	KP-24077
Status	Closed
Created	2018-06-14 11:01
Updated	2018-06-26 18:54
Linked Issues	-
Type Hits	Rikkekahtlus
CU cause	Mobile network coverage
Deadline	2018-06-17 11:06
Priority Date	High
EIC code	38ZEE-00003958-8
Type of customer contacts	-
Issue Type	-
Satellite ID	SLA
Technician's visit	Yes
CU solution	Antenna installation

METER INFORMATION	
Meter ID	152688
Metering point	772721
Serial Number	19959626
Meter created at	2014-03-20 15:08
Meter updated at	2013-12-18 10:11
Model	E35C(GPRS)+E350
Remotely readable	1
EIC Code	38ZEE-00003958-8

Figure 6-3: Features extraction example

6.3. Case Representation

The goal in developing a case-based decision support system is to create a knowledge repository that contains problem-solving experiences for the technician that helps them to solve problem in a very short time.

The knowledge repository, called “case library” in a CBR system, will contain the experience of the technician in the solving problems tasks. For this reason, the case library has to be easy to create, maintain and update.

The first step in the development of a CBR system is to decide the method by which the problem-solving experience is going to be represented in the system. The proposed representation model is based on a help-desk system, where it is usual to use textual attributes. My proposed case representation is based on HOMER help-desk case. [13]

The main features of this help-desk system are:

- The domain knowledge is represented in a flat list of attribute values to use textual attributes and base the system in a question-answer model. In this case, the modeling effort is quite low.

- Similarity calculation in such systems is based on surface features.

- To answer the questions, it is needed to create a decision tree by hand. With this option, the system avoids redundant information in a question-answer based textual approach.

Tickets have textual and numeral values, it is needed a case modeling based on textual features. In this type of systems, the effort during the knowledge acquisition step is higher but the results during the use and maintenance of the system are better.

6.3.1. Case modeling

The proposed case representation in the case library based on a help-desk operator models the data in Failures, symptoms and Solutions as represented in Figure 6-4.

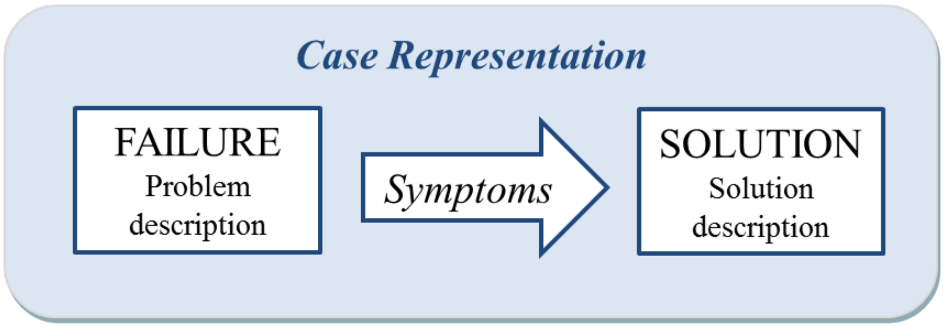

Figure 6-4: Case modeling

1) Failure: Information about the problem. In this section, details about the problem will be explained: cause, meter identifier, location and the description of the problem.

2) Symptoms: Information about the status of the meter relating to the problem. In this section, information about the meter (model, metering point and serial number) is also included.

3) Solution: Information about how the problem was solved. In this section, information about the solution, technician visit, linked issues and the description about how to solve the problem will be included.

6.3.2. Proposed case representation

An example of the proposed case representation is described in Table 6-3.

Table 6-3: Example of Case Representation

Case representation: Ticket 1	
FAILURE	- **EIC code:** 38ZEE-00003958-8 - **Cause:** Mobile network coverage - **Problem description:** - **Date:** 2018-06-14 11:01 - **Priority Date:** High
SYMPTOMS	- **Meter ID:** 152688 - **Metering point ID:** 772721 - **Serial Number:** 19959626 - **Model:** E35C(GPRS)+E350 - **Remotely readable:** 1
SOLUTION	- **Solution:** Antenna Installation - **Technician visit:** Yes - **Linked issues: -** - **Status:** Closed - **Solution description: -**

6.4. Case Library

One of the most important parts in a CBR system is the case library, where all the knowledge and information are stored in order to use them in the future. After decide the case representation, it is possible to build the case library.

In this thesis work, the case library was built manually using *Microsoft Excel 2010.* Extracting the features from the different files available at the beginning of the project, a repository with all the cases were created and saved with name “CaseLibrary.csv”. As it was described in the previous section 5.3, the information to create the cases were chose from the tickets, events/alarms, meter information and communication quality using the meter identifier to correlate the different files.

6.5. CBR system: develop of the CBR system using *myCBR Workbench*

To develop the CBR system proposed to this thesis work, the tool called *myCBR* was used. The *myCBR Workbench* was designed with improved communication between the system and the user, and support the user during the development phase of a CBR project; this includes the generation of case representations, the definition of similarity measures, the testing of retrieval and use of explanation functionality.

First of all, a project in *myCBR Workbench* was created. In this project, it is possible to build a case library, module the cases, adapt the similarity measures, and, at the end, use the Retrieval module of *myCBR* to analyze the results of the CBR system.

6.5.1. Import Case Library

The starting point of many CBR projects is the collection of initial case data. The existence of at least some case examples is usually a precondition for modeling an accurate case representation and corresponding similarity measures. For this reason, the first step in developing the CBR system is to import the existing data in *myCBR.*

The tool *myCBR* offers the possibility to generate cases from text, generate attribute description from LOD of import cases from CSV. As it was explained above, the Case Library was built in a CSV file in order to collect all the data from the available sources.

Once the Case Library was imported in *myCBR*, cases are represented as concepts and attributes. A concept is the function and attributes the features of every concept. The values and the types of the attributes are automatically filled when the case library is imported. If there are any missing data, the user can choose between some special values to add in these gaps: *_unknown_* or *_undefined_*.

Depending of the values defined in the case library, the type of the attribute is automatically chosen in *myCBR*. The type can be: *Boolean, concept, date, double, float, integer, interval, string* and *symbolic*. In the proposed case library the defined attributes are strings or symbolic, as it shown in Table 6-4.

Table 6-4: Attribute Types in myCBR

Attribute name	Attribute Type
Cause	*Symbolic*
Solution	*Symbolic*
Created	*String*
Deadline	*String*
EIC code	*String*
Issue Type	*Symbolic*
Key	*String*
Linked Issues	*String*
Meter created at	*String*
Meter ID	*String*
Meter updated at	*String*
Metering point ID	*String*
Model	*Symbolic*
Priority Date	*Symbolic*
Remotely readable	*Symbolic*
Satellite ID	*Symbolic*
Serial number	*String*
Status	*Symbolic*
Technician´s visit	*Symbolic*
Type Hits	*Symbolic*
Type of customer contacts	*Symbolic*
Updated	*String*

In Figure 6-5, the case library is show and also how looks like a ticket in the tool. For every ticket, it is possible to see all the features selected in section 5.3:

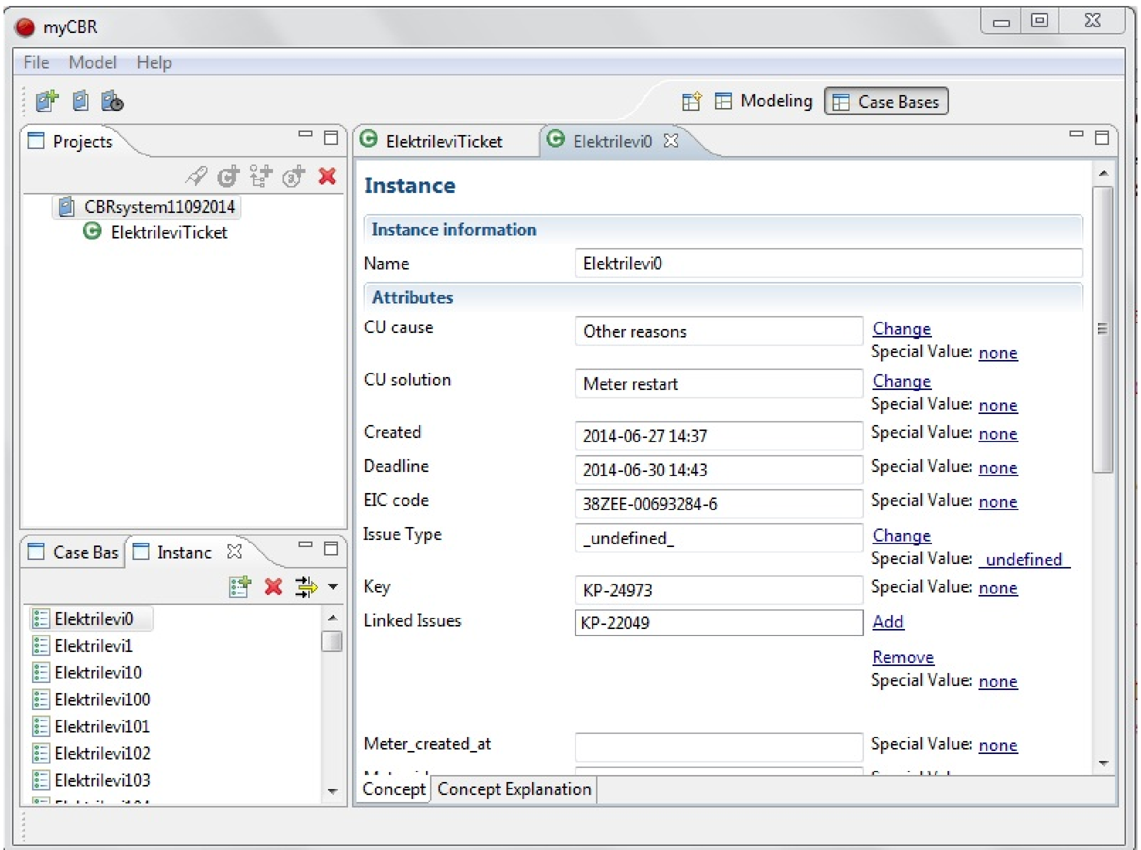

Figure 6-5: Case representation in myCBR

6.5.2. Model Similarity Measures

After having generated the case representation in *myCBR*, the next task for creating the CBR system is the definition of an appropriate similarity measure. *myCBR* follows the local-global approach which divides the similarity definition into a set of local similarity measures for each attribute, a set of attribute weights, and a global similarity measure for calculating the final similarity value.

Local Similarity Measures

The local similarity measure method depends on the type of attribute and values that this attribute have. *myCBR* defines different similarity approach for the value type and it is possible to model according to the data available. As it was described in the previous section, for the imported case library there are only two types of attributes: strings attributes and symbolic attributes. String values are not interested in the system for now. To define the local similarity measure only those symbolic attributes are considered.

For the proposed CBR system, the symbolic attributes have been selected to compare cases. For modeling symbolic attributes, several possibilities to model the similarity are supported. The most common is the definition of a similarity table where it is possible to define the similarity between all the pairwise value combinations and enumerate explicitly. It can be symmetric or asymmetric. The second option is the arrangement of symbols in taxonomy by using comfortable drag and drop functionality.

In the following Table 6-5 the considered attributes have been shown and also the similarity measure type.

Table 6-5: Local Similarity Measures in myCBR

Attribute name	Similarity Measure*	Type**
Cause	SimCause	SymbolFct
Issue Type	SimIssueType	SymbolFct
Model	SimModel	SymbolFct
Priority Date	SimPriority	SymbolFct
Remotely readable	SimReadable	SymbolFct
Satellite ID	SimSatellite	SymbolFct
Status	SimStatus	SymbolFct
Technician´s visit	SimVisit	SymbolFct
Type Hits	SimTypeHits	SymbolFct
Type of customer contacts	SimCustomer	SymbolFct

* *Similarity Measure are described in Appendix A*

***All the symbolic attributes have the type SymbolFct which defines a symmetric similarity table between the pair of values.*

Set of attributes weights

After modeling each local similarity measure, the next step is to define the weight of the attributes in the system. Giving different weight, the system will consider ones attributes more important than other. The set of attributes weights can be changed depending of the requirement of the system. The importance of each attribute should be discussed between the developer and the field technician.

For this system, the most important attributes are those which involve the problem description and the symptoms of the case. The attributes considered as false are not included in the global similarity measure because they are strings or can´t offer relevant information during the comparative. An example of set attributes weights is presented in Figure 6-6.

ElektrileviTicket | SimilarityElektrilevi

Type (•) Weighted Sum ( ) Euclidean ( ) Minimum ( ) Maximum

Attribute	Discriminant	Weight	SMF
CU cause	true	60.0	simCause
CU solution	true	20.0	simSolution
Created	false	0.0	default function
Deadline	false	0.0	default function
EIC code	false	0.0	default function
Issue Type	true	1.0	simIssue Type
Key	false	0.0	default function
Linked Issues	false	0.0	default function
Meter_created_at	false	0.0	default function
Meter_id	false	0.0	default function
Meter_updated...	false	0.0	default function
Metering_point...	false	0.0	default function
Model	true	1.0	simModel
Priority Date	true	1.0	simPriority
Remotely_read...	true	5.0	simReadable
Satellite ID	true	1.0	simSatelite
Serial_number	false	0.0	default function
Status	true	0.0	simStatus
Technician's vi...	true	5.0	simVisit
Type Hits	true	1.0	simTypeHits
Type of custo...	true	1.0	simCustomer
Updated	false	0.0	default function

Figure 6-6: Example of weights

Global Similarity Measures

The final step of the definition of the knowledge in the CBR system is to define the Global Similarity Measure, which will define the similarity between the matched cases during the retrieval function. Using *myCBR* it is possible to define different Global Similarity Measure to compare between them. The options for the Global Similarity Measures are:

1) *Weighted Sum:* The global similarity will be calculated following the equation:

$$Sim(a, b) = \sum w_i \, . \, sim_i(a_i,b_i)$$

Here, sim_i and w_i denote the local similarity measure and the weight of attribute i, and *Sim* represents the global similarity measure.

2) *Euclidean:* The distance between vectors *a* and *b* is defined as follows:

$$Sim(a,b)=\sqrt{\sum_i^n (x_i - y_i)^2}$$

In other words, Euclidean distance is the square root of the sum of squared differences between corresponding elements of the two vectors.

3) *Maximum of the local similarity measures:* The global similarity will be calculated following the equation:

$$Sim = max(weight * local\_sim)$$

4) *Minimum of the local similarity measures:* The global similarity will be calculated following the equation:

$$Sim = min(weight * local\_sim)$$

7. EVALUATION

Using the CBR tool *myCBR*, the thesis work presents a retrieval function for the proposed CBR system. This chapter explains the evaluation of the case library and the retrieval function.

7.1. Retrieval Functionality in myCBR

Once the case library was imported and the similarities measures were defined in the CBR tool called *myCBR*, it is time to use the retrieval functionality of this tool in order to test the correct definition of the proposed system. The definition of an optimal similarity measure is often a difficult and tricky task that requires repeatedly testing and fine tuning. For this purpose, *myCBR* includes a graphical user interface for performing retrievals and for analyzing the corresponding results.

Therefore, in order to test the resulting prototype of system, the *Retrieval* functionality in *myCBR Workbench* was used. As it was mentioned, during the development of the CBR system a case library of 547 cases was used. For the evaluation step of the proposed CBR system, it will be defined a small case library of 50 cases for training and other small case library of 50 cases for testing.

The *Retrieval* functionality in *myCBR Workbench* selects the similar tickets and presents the similarity when a new ticket is matched. This functionality helps to test and adjust the similarities measures in order to get the best results according to the specifications of the proposed system. To use the similarity-based retrieval in *myCBR*, the user has to select the value of the different attributes from the vocabulary of the case library.

When the user inserts a new ticket in *myCBR*, the tool compares the value of the attributes with the existing tickets and gets the similar tickets stored in the CBR system. In addition, the tool presents the similarity between the new ticket and all the existing tickets. As it was explained before, the similarity value depends on the type of the Global Similarity Measure matched during the development of the CBR system.

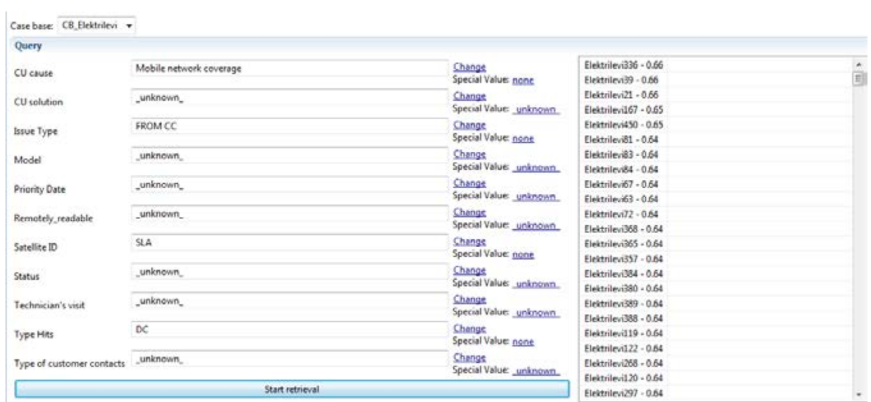

Figure 7-1: Retrieval Example

In Figure 7-1, an example of the *Retrieval* functionality is presented: Cause selected is “Mobile Network Coverage”, Issue Type is “FROM CC”, Satellite ID is “SLA” and Type Hits is “DC”. The better results of this searching are 0.66 and 0.65, but *myCBR* also presents other tickets.

7.2. Evaluation of the CBR system

The evaluation of the CBR system has two different part: first of all, the training part where the CBR system was prepared choosing an specific case library of 50 cases and the similarity measures were selected; and, secondly, the testing part, where the training case library was tested with a case library of other different 50 cases of the initial case library of 457 cases. After this evaluation, it was possible to analyze the accuracy of the CBR system according to different requirements.

7.2.1. Training the CBR system

To train the CBR system in *myCBR* a selection of 50 cases from the initial case library of 457 cases has been carried out. As it was explained before, one of the most important problems in this thesis work was the missing data; in order to use the maximum data available, the 50 cases used for this training part were selected choosing those cases where the maximum fields were filled in.

The other important step during the training part of the evaluation is the definition of the attributes and the similarity measures that will be used to test the CBR system.

For the evaluation, only the features regarding the problems have been selected, but the meter information has not been included because most of the cases don't have this information filled in. The selected features of the case library used for the evaluation are shown in Table 7-1. In *myCBR*, the features are defined as attributes and one of the most important steps during the CBR system development was the definition of the Local Similarity Measures for each attribute.

Local Similarity Measures

As it was explained before, each attribute has a Local Similarity Measures that has been selected according to the values and the types of values of the data. All the attributes are symbolic values and the local similarity measures have been declared following a symmetric similarity table for each pair of values. The attributes and similarities function used for testing the CBR system in *myCBR* are shown in Table 7-1.

Table 7-1: Selected features for evaluation

Attribute name	Local Similarity Measure*
Cause	SimCause
Issue Type	SimIssueType
Priority Date	SimPriority
Satellite ID	SimSatellite
Technician´s visit	SimVisit
Type Hits	SimTypeHits
Type of customer contacts	SimCustomer

*For more information about the Local Similarity Measures see Appendix A.

Set of attributes weights

To define the set of attributes weights, a table from most important features until the less has been defined. At the beginning, the selected features were 10, so from 10 to 1, each attribute has a weight. The order of importance and the weights defined of each attribute are shown in Table 7-2.

Table 7-2: Set of weight for evaluation

Attribute name	Weight
Cause	**10**
Type Hits	**9**
Issue Type	**8**
Satellite ID	**7**
Priority Date	**6**
Type of customer contacts	**5**
Technician´s visit	**4**
Remotely readable	**3**
Model	**2**
Status	**1**

Global Similarity Measures

During the evaluation, two different types of global similarity measures have been used: the weight sum and the Euclidean. The global similarity measure allows defining the similarity between the new case and the previous cases that the CBR system shows during the retrieval functionality.

Comparing between both global similarities measures, the second measures used, the Euclidean permits to get higher similarity between the matched cases. Therefore, it is the global similarity measures used during the testing.

7.2.2. Testing the CBR system

To test the CBR system, it has been used 50 cases from the initial case library of 457 cases. The testing case library has different cases than the training case library. Using the testing case library and the Retrieval functionality in *myCBR*, it is possible to define the accuracy of the proposed CBR system.

After importing the training case library in *myCBR* and selecting the similarity measure of the CBR system, the evaluation started. The solutions of all the cases of the testing case library were compared with the solution that the CBR system proposed according to the training case library. For each new case, the selected features were matched manually in the *Retrieval* Functionality in *myCBR* in order to get the most similar cases stored in the system.

For each tested case, if the solution retrieved is the same as the current solution, the test is considered as correct and matched with ´1´. If it is not the same, the test is fail and matched with ´0´.

During the testing of the CBR system, three options were considered:

- *K=1:* Matching the first solution given by the CBR system.
- *K=2*: Choosing the best solution between the first and the second solutions given by the CBR system. The best solution is considered when it matches with the solution of the studied case.
- *K=3*: Choosing the best solution between the three firsts options given by the CBR system.

7.2.3. Results: accuracy of the CBR system

The accuracy of the CBR system is calculated with the following equation:

$$Accuracy\ (\%) = \left(\frac{Number\ of\ correct\ solutions}{50}\right) \cdot 100$$

Considering the number of correct solutions and the whole case library of 50 cases, the result of the evaluation for the three considered options during this testing are shown in Table 7-3.

Table 7-3: Accuracy of the CBR system

K	# Correct	Accuracy
1	39	78 %
2	42	84 %
3	42	84 %

The best accuracy of the system is considering the first and the second solutions given by the CBR system and this accuracy doesn't change if the user decides to consider the third solution. These results mean that in a CBR system, the first option is not always the best one. It is responsibility of the user, in this case, of the technician, to choose the best solution considering other parameters.

8. CONCLUSIONS AND FUTURE WORK

After the evaluation of the proposed CBR system, this chapter presents the conclusions and the future work of this thesis work.

8.1. Conclusions

Decision support systems help technicians and managers to make decisions in specific situations using information from documents, raw data, knowledge and other sources. In the electricity network, the main focus of the DSS is to propose solutions to the technicians and help them to make decisions when a new problem is reported in the meters. Using the previous data of the network, this thesis work proposes a DSS system based on CBR techniques. This CBR system will help technicians during fault diagnosis and maintenance of the electricity network reusing other previous problem. Thus, the electricity company could improve the quality of service because the technician it will save time to solve typical problems in the electricity networks. Moreover, the company could give more effective solution when a problem is reported. Therefore, the proposed system will help to save time as well as money in such domain.

During the development of the CBR system, the following tasks have been done problem domain analysis, data analysis, knowledge acquisition, feature extraction, case formulation and development of a case library for the CBR system. In addition, in order to develop the retrieval step of the CBR cycle proposed by Agnar and Plaza, the thesis work proposes to use the CBR tool called *myCBR.*

The problem domain analysis finished with the definition of the problems in the electricity network and the possible solutions for each kind of problem. The distribution of problems and the relation between the causes and the solutions are also presented.

During the data analysis, the available data are cleaned in order to select relevant information from the electricity network. However, the main problem in this thesis work is the missing data and incomplete features of the data. On the other hand, the data are from different sources and one of the difficult steps of this analysis was the correlation of the data from these files.

After the analysis, it is time to create the proposed case library for the CBR system. In this case library, 547 cases were considered. The proposed case library obtained an accuracy of 84%, but it could include using more features that help during the definition of the problem.

The result of the evaluation shows that in a CBR system the first solution is not always the best, the technician need more parameters to choose the solution that solves the reported problem. In the CBR technique, the system proposes several solutions with different probabilities and the user has to decide which of these solutions could be the best for the problem.

8.2. Future Work

This thesis work established the basics of the proposed decision support system, but for the future, it will be needed to develop the system according to the requirements of the electricity company.

For the future, the following cases of study are proposed:

- Analyze data from other nodes of the electricity network: For this thesis work, the system only uses a small section of the electricity network. The case library could improve including more features from the network.
- Analyze each cause to understand the features that are more important in each case. This analysis will help to define new similarity measures and a different set of attribute weight for each cause.
- Increase the selected features in the proposed case library: including other features like the communication quality or type events and alarms in the case library, the field technician will be able to understand the status of the system when a problem is reported.
- Develop a feature extraction algorithm that automates the building of the case library and updates the case library when the new features will be defined.
- Develop the CBR cycle in the proposed CBR system including retrieval, reuse, revise and retain steps in order to propose a GUI to the company.
- Integrate the CBR system in the electricity network.
- Compare the proposed CBR system with other reasoning systems.
- Calculate the accuracy of the CBR system when it is use with other reasoning system like decision trees and random forest. The accuracy could increase including other reasoning techniques in the system.

References

[1] **Ian Watson** (1997); *"Applying Case-Based Reasoning: Techniques for Enterprise Systems"*, Morgan Kaufmann, CA, USA, 1997.

[2] **A. Aamodt, E. Plaza** (1994); "*Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches*", AI Communications. IOS Press, Vol. 7: 1, pp. 39-59.

[3] **S. Begum, M. U.Ahmed, P. Funk, N. Xiong and M. Folke** (July, 2011); *"Case-Based Reasoning Systems in the Health Sciences: A Survey of Recent Trends and Developments"*, in IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—Part C: Applications and Reviews, Vol. 41, no. 4

[4] **S Begum, MU Ahmed, P Funk;** "*ECG sensor signal analysis to represent cases in a case-based stress diagnosis system"*, 10th IEEE International Conference on Information Technology and Applications in Biomedicine (ITAB), Corfu, Greece

[5] **G. Valente and A. Rigallo**; "*Using Case-Based Reasoning to support Operational Knowledge Management"*, Università di Torino, Italy

[6] **Stuart J. Russell and Peter Norvig** (1995)**; "***Artificial Intelligence a modern approach"*, Series in Artificial Intelligence, Prentice Hall, Englewood Cliffs, Chapter 17

[7] **Keen, Peter** (1980); *"Decision support systems: a research perspective"*, Cambridge, Mass: Center for Information Systems Research, Afred P. Sloan School of Management. http://hdl.handle.net/1721.1/47172

[8] **Gorry, A. G. & Morton, S. S.** (1971); "*Framework for Management Information Systems*", Sloan Management Review, 13.

[9] **Keen, G. P. & Morton, M. S.** (1978); *"Decision Support Systems: An Organizational Perspective*", Reading, MA: Addison-Wesley.

[10] **Bonczek, R. H., Holsapple, C. W. & Whinston, A. B**. (1980); *"The Evolving Roles of Models in Decision Support Systems"*, Decision Sciences, 11(2).

[11] **Turban, E**. (1995); *"Decision Support and Expert Systems: Management Support Systems*", N J: Englewood Cliffs.

[12] **Turban, E., Aronson, J. E., Liang, T. P. & Sharda, R.** (2007); *"Decision Support and Business Intelligence Systems"*, Noida: Pearson Education, Inc.

[13] **Mehmet H. Gökera, T. Roth-Berghofer** (1999); *"The development and utilization of the case-based help-desk support system HOMER"*, Engineering Applications of Artificial Intelligence 12 (1999), pp. 665-680

[14] **T. Roth-Berghofer** (2004); *"Learning for HOMER, a Case Based Help-Desk Support System"*, G. Melnik and H. Holz (Eds.): LSO 2004, LNCS 3096, pp. 88–97

[15] **C. Melchiors and Liane M. R. Tarouco** (1999); *"Fault Management in Computer Networks Using Case-Based Reasoning: DUMBO System",* PPGC, Universidade Federal do Rio Grande do Sul. K.-D. Althoff, R. Bergmann, L.K. Branting (Eds.), ICCBR-99, LNAI 1650, pp. 510-524

[16] **N.Amani, M. Fathi and M. Dehghan** (2005); *"A case-based reasoning method for alarm filtering and Correlation in telecommunication networks"*, CCECE / CCGEI, Saskatoon, May 2005, pp. 2176-2180

[17] **K. Bach, Klaus-Dieter Althoff, R.Newó and Armin Stahl** (2011); *"A Case-Based Reasoning Approach for Providing Machine Diagnosis from Service Reports".* A. Ram and N. Wiratunga (Eds.): ICCBR 2011, LNAI 6880, pp. 363–377

[18] **Yvh Foong David Law, Sew Bun Foong And Shee Eng Jeremiah Kwan** (1997)**;** *"An Integrated Case-Based Reasoning Approach for Intelligent Help Desk Fault Management"*, National University of Singapore, Expert Systems With Applications, Vol. 13, No. 4, pp. 265-274

[19] **Richard Heider** (1996); *"Troubleshooting CFM 56-3 Engines for the Boeing 737 Using CBR and Data-Mining"*, Lecture Notes in Computer Science, vol. 1168, pp. 512-523

[20] **Gunnar Ingi Ómarsson** (2010); *"Knowledge Management in an IT-Help Desk environment"*, University of Skövde

[21] **A.R. Farjadi Tehrani and F. Z. Mustafa Mohamed** (2011); *"A CBR-based Approach to ITIL-based Service Desk"*, Journal of Emerging Trends in Computing and Information Sciences , VOL. 2, NO. 10, pp. 476- 484

[22] **Azeb Bekele Eshete** (2009)**; "***Integrated Case Based and Rule Based Reasoning for Decision Support"*, Norwegian University of Science and Technology

[23] **Andreas Hanemann; "***A Hybrid Rule-Based/Case-Based Reasoning Approach for Service Fault Diagnosis"*, Munich Network Management Team

[24] **J. Nordlund and H. Schäfer** (2006); *"Case-Based Reasoning in a Support system"*. Master's Thesis in Computing Science, Umeå University

[25] **W. Cheetham and I. Watson** (2006); *"Fielded applications of case-based reasoning"*, The Knowledge Engineering Review, Vol. 20:3, pp. 321–323

[26] **A. Tariq and K. Rafi** (2012); *"Intelligent Decision Support Systems- A Framework"*, Information and Knowledge Management, ISSN 2224-5758 (Paper) ISSN 2224-896X (Online), Vol 2, No.6

[27] **David W. Aha and I. Watson** (2001); *" Case-Based Reasoning Research and Development"*, 4th International Conference on Case-Based Reasoning, ICCBR 2001 Vancouver, BC, 2001

[28] **I. Watson** (1999); *"Case-based reasoning is a methodology not a technology"*, Knowledge-Based Systems 12 (1999) 303–308

[29] **Kolonder, J. L.** (1993); *"Case-based reasoning"*, San Mateo, CA : Morgan Kaufmann

[30] **Julie Main, Tharam Dillon and Simon Shiu** (2001); *"A Tutorial on Case-Based Reasoning"*, Soft Computing in Case Based Reasoning, (Eds.) Sankar K Pal, Tharam Dillon and Daniel Yeung, Springer-Verlag (London) Ltd, 2001, pp.1-28

[31] **Richter, M.M.** (1998); *Introduction*. In: Lenz, M., Bartsch-Sp□orl, B., Burkhard, H.D., Wess, S. (eds.) Case-Based Reasoning Technology - From Foundations to Applications. LNAI 1400, Springer-Verlag, Berlin

[32] **Bergmann, R. and Stahl, A.** (1998); *"Similarity Measures for Object-Oriented Case Representations"*, Proceedings of the 4th European Workshop on Case-Based Reasoning (EWCBR'98)

[33] **Laura Zilles** (2009); *"myCBR Tutorial",* myCBR Project, April 11, 2009

[34] **K.Bach and K.-D. Althoff** (2012); *"Developing Case-Based Reasoning Applications Using myCBR 3",* Case-based Reasoning in Research and Development, Proceedings of the 20th International Conference on Case-Based Reasoning (ICCBR-12), Springer, LNAI 7466, pp. 17-31, Eds: Watson, Ian and Diaz Agudo, Belen, September 2012

[35] **K. Bach, C. Sauer, K.-D. Althoff and T.Roth-Berghofer**; *"Knowledge Modeling with the Open Source Tool myCBR"*

[36] **A. Stahl and T. Roth-Berghofer** (2008); *"Rapid prototyping of CBR applications with the open source tool myCBR",*Ralph Bergmann and Klaus-Dieter Altho, editors, Advances in Case-Based Reasoning. Springer Verlag, 2008.

[37] **T. Roth-Berghofer, C.S. Sauer, J. Recio Garcia, K. Bach, K.-D. Althoff, and B. Diaz Agudo** (2008); "*Building case-based reasoning applications with myCBR and COLIBRI studio*", Proceedings of the UKCBR 2012 Workshop

[38] **A. Atanassov, L. Antonov** (2012); "*Comparative analysis of Case Based Reasoning software frameworks jCOLIBRI and myCBR"*, Journal of the University of Chemical Technology and Metallurgy, 47, 1, 2012, 83-90

[39] **D. Bahls and T.Roth-Berghofer**; "*Explanation Support for the Case-Based Reasoning Tool myCBR"*,German Research Center for Artificial Intelligence DFKI GmbH

[40] **myCBR:** http://www.mycbr-project.net

[41] **jCOLIBRI:** http://gaia.fdi.ucm.es/research/colibri

APPENDIX

A.Local Similarity Measures in *myCBR*

In the proposed CBR system using the tool *myCBR*, the local similarity measures proposed for the symbolic attributes are defined by symmetric similarity tables between the pair of values. The following functions present the values of each attribute and the similarities between the values used during the developing of the CBR system.

1) SimCause

The Local Similarity Measure called *SimCause* defines the symmetry between the causes of the problems. The symbolic attribute *Cause* can adopt the following values: Main switch, Meter Failure, Mobile Network Coverage, Noise or Other reasons:

Table A - 1: SimCause

Cause	*Main Switch*	*Meter Failure*	*Mobile NW*	*Noise*	*Other reasons*
Main Switch	1	0	0	0	0
Meter Failure	0	1	0	0	0
Mobile NW	0	0	1	0	0
Noise	0	0	0	1	0
Other reasons	0	0	0	0	1

2) SimIssueType

The Local Similarity Measure called *SimIssueType* defines the symmetry between the issues types. The symbolic attribute *Issue Type* can adopt the following values: Client, Internal Ericsson or From CC:

Table A - 2: SimIssueType

Issue Type	*Client*	*Internal Ericsson*	*From CC*
Client	1	0	0
Internal Ericsson	0	1	0
From CC	0	0	1

3) SimModel

The Local Similarity Measure called *SimModel* defines the symmetry between the models of the meter. There are 3 different models of meters: Model A, Model B or Model C:

Table A - 3: SimModel

Model	*Model A*	*Model B*	*Model C*
Model A	1	0	0
Model B	0	1	0
Model C	0	0	1

4) SimPriority

The Local Similarity Measure called *SimPriority* defines the symmetry of the priority date. The *Priority Date* attribute can be: High, Medium or Low:

Table A - 4: SimPriority

Priority Date	*High*	*Medium*	*Low*
High	1	0	0
Medium	0	1	0
Low	0	0	1

5) SimReadable

The symbolic attribute *Remotely Readable* can be YES (1) or NO (0):

Table A - 5: SimReadable

Remotely Readable	*1 (YES)*	*0 (NO)*
1 (YES)	1	0
0 (NO)	0	1

6) SimSatellite

The Local Similarity Measure called *SimSatellite* defines the symmetry between the types of the satellites in the network. The symbolic attribute *Satellite ID* can adopt the following values: PRE-SAT, SAT or SLA:

Table A - 6: SimSatellite

Satellite ID	*PRE-SAT*	*SAT*	*SLA*
PRE-SAT	1	0	0
SAT	0	1	0
SLA	0	0	1

7) SimStatus

The symbolic attribute *Status* can be Open or Closed:

Table A - 7: SimStatus

Status	*Open*	*Closed*
Open	1	0
Closed	0	1

8) SimVisit

The symbolic attribute *Technician visit* can be YES (1) or NO (0):

Table A - 8: SimVisit

Technician visit	*1 (YES)*	*0 (NO)*
1 (YES)	1	0
0 (NO)	0	1

9) SimTypeHits

The Local Similarity Measure called *SimTypeHits* defines the symmetry between the types of hits. The symbolic attribute *Type Hits* can adopt the following values: DC, PLC or Fault:

Table A - 9: SimTypeHits

Type Hits	*DC*	*PLC*	*Fault*
DC	1	0	0
PLC	0	1	0

Fault	0	0	1

10) SimCustomer

The Local Similarity Measure called *SimCustomer* defines the symmetry between the types of the satellites in the network. The symbolic attribute *Customer* can adopt the following values: PLC, Info, Meter display, Malfunction or Main fuse:

Table A - 10: SimCustomer

Customer	*PLC*	*Info*	*Meter display*	*Malfunction*	*Main fuse*
PLC	1	0	0	0	0
Info	0	1	0	0	0
Meter display	0	0	1	0	0
Malfunction	0	0	0	1	0
Main Fuse	0	0	0	0	1

APPENDIX

B. Introducción

Hoy en día, la industria de las telecomunicaciones es un mercado muy amplio y muy accesible para empresas de todos los ámbitos. Con el objetivo de ser los primeros en el mercado, empresas en todo el mundo quieren mejorar y actualizar sus sistemas.

Debido al rápido avance de la tecnología, las empresas necesitan actualizar sus sistemas continuamente para adaptarse a las necesidades del mercado y poder competir con el resto de empresas. Actualmente, con el fin de obtener mayores beneficios y de ahorrar tiempo y dinero, las empresas están incluyendo sistemas de diagnóstico y monitorización en sus redes para la solución de problemas en el menor tiempo posible, lo que supone que también necesitan actualizar su software y adaptarse a las nuevas tecnologías.

En el ámbito de la Inteligencia Artificial (AI, Artificial Intelligence en inglés), se utilizan técnicas que ayudan en la solución de problemas y el apoyo en la toma de decisiones. AI es un término que se ocupa de entender la inteligencia humana. Por lo tanto, para el desarrollo de sistemas basados en la inteligencia humana, en primer lugar, es necesario estudiar el comportamiento humano para después desarrollar sistemas de acuerdo a estas bases. En otras palabras, un sistema basado en la inteligencia humana deben poseer las siguientes capacidades: procesamiento del lenguaje natural, representación del conocimiento, razonamiento automático y aprendizaje automático.

Este Proyecto de Fin de Carrera se centra en el diseño de un sistema de soporte a decisiones (DSS, Decision Support System en inglés) utilizando técnicas basadas en la Inteligencia Artificial. Los sistemas de soporte a decisiones se utilizan para tomar decisiones y resolver problemas ofreciendo soluciones en la mayor brevedad posible. Los expertos se basan en la experiencia previa y el conocimiento en un determinado ámbito para poder tomar las decisiones, por tanto, un sistema que ayude en las decisiones debe basarse también en la experiencia y debe comportarse de acuerdo al razonamiento humano para resolver los nuevos problemas que puedan surgir. En AI, los sistemas que utilizan el conocimiento y experiencia previa y que se basan en el razonamiento humano pueden implementarse utilizando un algoritmo llamado razonamiento basado en casos (CBR, Case-Based Reasoning en inglés).

B.1. Motivación

En los centros de operaciones de red es muy importante tener un buen sistema de detección de problemas en tiempo real que permita ofrecer soluciones de forma inmediata ante cualquier problema que pueda surgir. Estos centros de operaciones de red ayudan a las empresas a administrar, controlar y monitorear sus redes. Cuando hablamos de centros de operaciones de red con un elevado número de usuarios es necesario tener un sistema capaz de buscar una solución rápida y eficaz para evitar que el sistema sufra grandes pérdidas, sobre todo cuando se trata de problemas que se pueden repetir habitualmente o que suponen un gran coste.

Hoy en día, la investigación se centra en sistemas de diagnóstico y mantenimiento basados en el comportamiento humano que ayuden en la toma de decisiones. Los sistemas deben ser capaces de controlar y detectar eventos en tiempo real, así como, resolver problemas en el menor tiempo posible para evitar grandes pérdidas a las empresas.

Los expertos deben resolver los problemas en las redes con la información almacenada en los centros de operaciones de red, pero a menudo no es tan fácil como pueden imaginar. Se necesitan otras ayudas para encontrar la causa real del problema y definir la mejor solución en este caso. Los sistemas de soporte a decisiones podrían ser un útil durante el proceso de recolección.

En este proyecto, se pretende desarrollar un sistema basado en el Razonamiento basado en casos (CBR) capaz de ofrecer soluciones en tiempo real a problemas que surjan durante el funcionamiento habitual de un centro de operaciones de red basándose en problemas similares, conocimiento previo y la experiencia.

El Razonamiento basado en casos es un algoritmo que propone la reutilización de problemas pasados, o casos, para ofrecer una solución a un nuevo problema. Para ello, es necesario buscar las similitudes del nuevo problema con los de la base de datos de casos y reutilizar la solución que más nos convenga. En otras palabras, se basa en la recuperación y adaptación de las soluciones anteriores para el problema actual. El algoritmo diferencia cuatro fases: Retrieve, donde se compara el nuevo problema con los casos previos y se analiza la similitud entre ellos; Reuse, donde se reutiliza la solución del problema más similar; Revise, para revisar que la solución es válida para el nuevo problema; y Retain, donde se almacena el problema resuelto y su solución para futuros usos.

La idea fundamental del proyecto es analizar los datos y estudiar las características de una red para crear una base de datos que nos permita desarrollar un sistema de soporte a decisiones basado en la técnica del CBR. Con ello, el centro de operaciones de red podrá detectar el momento en el que se produce el fallo y ofrecer una solución en el menor tiempo posible. Por tanto, será necesario desarrollar la forma de extraer las características más importantes del sistema y definir las similitudes entre los problemas.

B.2. Antecedentes

La red de distribución de la energía eléctrica es la parte del sistema de suministro eléctrico cuya función es el suministro de energía desde la subestación de distribución hasta los usuarios finales. Se lleva a cabo por los Operadores del Sistema de Distribución (Distribution System Operator, DSO, en inglés). Los DSOs, por tanto, tienen tanto la responsabilidad de entregar sin fallos la energía eléctrica de los proveedores a los clientes, como de mantener dichas redes de distribución.

Con el fin de mejorar la calidad de la atención al cliente, el DSO tiene definir dos objetivos diferentes: en primer lugar, la renovación de la red de distribución y, en segundo lugar, la instalación de un nuevo sistema de medición inteligente, en el cuál todos los medidores estarán monitorizados para poder ofertar el mejor precio de la energía a los clientes según los valores de consumo. Los medidores inteligentes ayudarán a la DSO en el seguimiento del mercado de la energía. El sistema de medición inteligente debe ser determinado por la forma en que se puede mejorar el servicio al cliente y el funcionamiento de la red al menor precio posible. Por otro lado, el tiempo necesario resolver los problemas en la red es a veces demasiado y por lo tanto cuesta demasiado; por esta razón, necesitan un sistema de soporte de decisiones capaces de ofrecer soluciones en el menor tiempo posible.

B.3. Objetivos

El objetivo de este proyecto es analizar los datos de la red eléctrica para proponer un sistema capaz de ofrecer soluciones a los problemas en la red eléctrica utilizando el Razonamiento basado en casos (CBR, en inglés, Case-Based Reasoning) para capturar información desde un centro de operaciones de red en tiempo real.

La principal idea es analizar los datos ofrecidos por la empresa de electricidad y capturar aquellos parámetros que nos permitan conocer cuándo hay un problema para crear un nuevo caso. El nuevo caso se tratará en el sistema a desarrollar buscando una solución a partir de casos similares que se hayan almacenado en la base de datos de casos previos, es decir, se quieren reutilizar soluciones anteriores y adaptarlas al nuevo caso de forma que se pueda ofrecer una nueva solución y así poder trabajar con predicciones de futuro cuando sea necesario.

Como objetivo final del proyecto, se quiere desarrollar un sistema que nos permita encontrar soluciones rápidas y con el menor coste posible para la empresa para que se utilice como parte del proceso de localización de problemas en red (Troubleshooting), es decir, durante el análisis, planteamiento y ejecución de un problema.

B.4. Metodología y plan de trabajo

Tras el análisis del problema y la definición de los objetivos del trabajo de tesis, se proponen las siguientes tareas para llevar a cabo el proyecto:

1. Comprender la red eléctrica del sistema actual

2. Análisis de los datos y de las necesidades del sistema de soporte de red, búsqueda y decisión de las funcionalidades que debería incluir el sistema a proponer

3. Búsqueda y análisis de sistemas similares actuales: se trata de conseguir la información sobre los programas que se utilizan actualmente, cómo capturan los datos y cómo se analizan

4. Proposición de un método en el ámbito de la Inteligencia Artificial: Razonamiento basado en casos

5. Diseño del prototipo con los resultados obtenidos en los estudios anteriores

6. Análisis de resultados y conclusiones

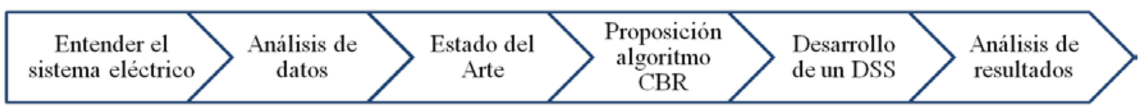

Figure B-1: Metodología

B.5. Estructura

Capítulo 1 introduce el problema y los objetivos de este Proyecto de Fin de Carrera.

Capítulo 2 describe los sistemas que actualmente se utilizan en el ámbito de la Inteligencia Artificial, así como en los sistemas de soporte a decisiones. Además, explica el algoritmo llamado Razonamiento basado en casos y ayuda a comprender los sistemas basados en este método. Este capítulo también incluye limitaciones y ventajas de los métodos.

Capítulo 3 presenta el estado del arte en los sistemas basados en CBR y los avances en estas técnicas.

Capítulo 4 explica el problema propuesto y describe la red eléctrica del sistema para el que se basa este proyecto.

Capítulo 5 describe el análisis de los datos y los distintos problemas y soluciones de la red eléctrica.

C**apítulo 6** presenta los pasos seguidos para llevar a cabo la base de datos y se explica el sistema propuesto.

Capítulo 7 explica la evaluación del sistema CBR y los resultados.

Capítulo 8 presenta la conclusión del proyecto y el trabajo futuro.

APPENDIX

C. Conclusiones y Trabajo Futuro

C.1. Conclusiones

Los sistemas de soporte a decisiones (DSS) pueden ayudar a los técnicos a tomar decisiones en determinadas situaciones utilizando información de distintas fuentes como pueden ser documentos, experiencia o casos previos. En este Proyecto de Fin de Carrera, la técnica propuesta para desarrollar un DSS fue el razonamiento basado en casos (CBR), algoritmo que propone una solución utilizando el conocimiento y la experiencia previa.

Para la red eléctrica propuesta, el principal objetivo es crear un DSS que pueda proponer soluciones a los técnicos y ayudarles a tomar decisiones cuando un nuevo problema se presenta en los medidores. Utilizando los datos anteriores de la red, este proyecto propone un sistema DSS basado en el algoritmo CBR. Este sistema CBR ayudar a los técnicos durante el diagnóstico de fallos y el mantenimiento de la red eléctrica recuperando información sobre otro problemas anteriores. Con este sistema, la compañía eléctrica podría mejorar la calidad del servicio debido a que los técnicos podrán resolver los problemas en menos tiempo. Además, la empresa podrá ofrecer soluciones menos costosas para la empresa cuando se notifica un problema. Como conclusión, la empresa podría ganar tiempo y dinero con el sistema propuesto.

Durante el desarrollo del sistema CBR, las siguientes tareas se han terminado: estudio de la red eléctrica, análisis de los datos recibidos, la adquisición de conocimientos, la extracción de características y parámetros importantes, la definición de la representación de los casos y el desarrollo de una base de datos para el sistema CBR. Además, con el fin de proponer la etapa de recuperación del ciclo CBR propuesto por Agnar y Plaza, el trabajo de tesis propone utilizar la herramienta llamada *myCBR*.

El estudio de la red eléctrica terminó con la definición de los problemas en la red eléctrica y las posibles soluciones para cada tipo de problema. También se presentaron la distribución de los problemas y la relación entre las causas y las soluciones.

Durante el análisis de los datos, se extrajeron los datos relevantes para nuestro sistema con el fin de seleccionar los parámetros característicos de la red eléctrica. Sin embargo, el problema principal que se ha encontrado durante la ejecución de este proyecto ha sido la falta de información debido a que los ficheros analizados contaban con datos incompletos y/o sin información. Por otro lado, también se ha encontrado que no existe correlación entre los datos disponibles, por lo que una de las partes más complicadas del proyecto ha sido el análisis y extracción de características de los problemas encontrados en los archivos.

Después del análisis, se ha creado una base de datos de los problemas de la red eléctrica específica para el sistema CBR. En esta base de datos, se han considerado 547 casos. La base de datos propuesta obtuvo una precisión del 84%, pero podría mejorar si se incluyen más características que ayuden a definir los problemas.

El resultado de la evaluación, determinó que en un sistema CBR la primera solución no siempre es la mejor, por tanto, el técnico necesita más parámetros para elegir la solución que resuelve el problema reportado. En la técnica CBR, el sistema propone varias soluciones con diferentes probabilidades y el usuario tiene que decidir cuál de estas soluciones podría ser la mejor para el problema, y este proyecto también presenta el mismo resultado.

C.2. Trabajo Futuro

Este Proyecto de Fin de Carrera establece los aspectos básicos del sistema de soporte a decisiones propuesto, pero para el futuro, será necesario desarrollar un sistema que pueda cumplir con todos los requisitos de la empresa de electricidad.

Para el futuro, se proponen los siguientes casos de estudio:

- Analizar los datos de otros nodos de la red eléctrica: Para este proyecto, el sistema sólo utiliza una pequeña sección de la red eléctrica. Analizando más datos de la red, será posible incluir más características en la base de datos.

- Analizar cada tipo de problema para entender las características que son más importantes en cada caso. Este análisis ayudará a definir nuevas medidas de similitud y otorgar la importancia adecuada a cada característica para cada uno de los problemas.

- Aumentar los parámetros de estudio en la base de datos: incluyendo otras características como la calidad de la comunicación o los tipos de eventos y alarmas, los técnicos serán capaz de entender el estado del sistema cuando se notifica un problema.

- Desarrollar un algoritmo de extracción de características que automatice la construcción y actualización de la base de datos, y que además, pueda incorporar nuevos datos cuando se definan las nuevas características.

- Desarrollar el ciclo completo del algoritmo CBR en el sistema propuesto, incluyendo la recuperación (Retrieve), reutilización (Reuse), revisión (Revise) y retención (Retain) de los casos con el fin de proponer una interfaz gráfica de usuario para la empresa. La herramienta propuesta es myCBR.

- Integrar el sistema CBR en la red eléctrica.

- Comparar el sistema CBR propuesto con otros sistemas de razonamiento.

- Calcular la precisión del sistema CBR cuando se utiliza con otro sistema de razonamiento como los árboles de decisión y los random forest. La precisión se incrementará incluyendo otras técnicas de razonamiento en el sistema.

APPENDIX

D.Presupuesto

1) **Ejecución Material**
   - Compra de ordenador personal (Software incluido) 2000 €
   - Material de oficina 200 €

2) **Honorarios Proyecto**
   - 1280 horas a 16€/hora 20480 €

3) **Material fungible**
   - Gastos de impresión 180 €
   - Encuadernación 50 €

4) **Subtotal del proyecto**
   - Subtotal Presupuesto 22910 €

5) **I.V.A aplicable**
   - 21 % Subtotal Presupuesto 4811,1 €

6) **Total presupuesto**

   - Total Presupuesto 27721,1 €

Madrid, Noviembre 2014
La Ingeniera Jefa de Proyecto

Fdo.: Berta Lorenzo Espinar
Ingeniera Superior de Telecomunicación

APPENDIX

E. Pliego de condiciones

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto, de un *SISTEMA DE SOPORTE A DECISIONES: CAPTURA E INTERCAMBIO DE CONOCIMIENTO EN UNA RED DE TELECOMUNICACIONES.* En lo que sigue, se supondrá que el proyecto ha sido encargado por una empresa cliente a una empresa consultora con la finalidad de realizar dicho sistema. Dicha empresa ha debido desarrollar una línea de investigación con objeto de elaborar el proyecto. Esta línea de investigación, junto con el posterior desarrollo de los programas está amparada por las condiciones particulares del siguiente pliego.

Supuesto que la utilización industrial de los métodos recogidos en el presente proyecto ha sido decidida por parte de la empresa cliente o de otras, la obra a realizar se regulará por las siguientes:

Condiciones generales

1. La modalidad de contratación será el concurso. La adjudicación se hará, por tanto, a la proposición más favorable sin atender exclusivamente al valor económico, dependiendo de las mayores garantías ofrecidas. La empresa que somete el proyecto a concurso se reserva el derecho a declararlo desierto.

2. El montaje y mecanización completa de los equipos que intervengan será realizado totalmente por la empresa licitadora.

3. En la oferta, se hará constar el precio total por el que se compromete a realizar la obra y el tanto por ciento de baja que supone este precio en relación con un importe límite si este se hubiera fijado.

4. La obra se realizará bajo la dirección técnica de un Ingeniero Superior de Telecomunicación, auxiliado por el número de Ingenieros Técnicos y Programadores que se estime preciso para el desarrollo de la misma.

5. Aparte del Ingeniero Director, el contratista tendrá derecho a contratar al resto del personal, pudiendo ceder esta prerrogativa a favor del Ingeniero Director, quien no estará obligado a aceptarla.

6. El contratista tiene derecho a sacar copias a su costa de los planos, pliego de condiciones y presupuestos. El Ingeniero autor del proyecto autorizará con su firma las copias solicitadas por el contratista después de confrontarlas.

7. Se abonará al contratista la obra que realmente ejecute con sujeción al proyecto que sirvió de base para la contratación, a las modificaciones autorizadas por la superioridad o a las órdenes que con arreglo a sus facultades le hayan comunicado por escrito al Ingeniero Director de obras siempre que dicha obra se haya ajustado a los preceptos de los pliegos de condiciones, con arreglo a los cuales, se harán las modificaciones y la valoración de las diversas unidades sin que el importe total pueda exceder de los presupuestos aprobados. Por consiguiente, el número de unidades que se consignan en el proyecto o en el presupuesto, no podrá servirle de fundamento para entablar reclamaciones de ninguna clase, salvo en los casos de rescisión.

8. Tanto en las certificaciones de obras como en la liquidación final, se abonarán los trabajos realizados por el contratista a los precios de ejecución material que figuran en el presupuesto para cada unidad de la obra.

9. Si excepcionalmente se hubiera ejecutado algún trabajo que no se ajustase a las condiciones de la contrata pero que sin embargo es admisible a juicio del Ingeniero Director de obras, se dará conocimiento a la Dirección, proponiendo a la vez la rebaja de precios que el Ingeniero estime justa y si la Dirección resolviera aceptar la obra, quedará el contratista obligado a conformarse con la rebaja acordada.

10. Cuando se juzgue necesario emplear materiales o ejecutar obras que no figuren en el presupuesto de la contrata, se evaluará su importe a los precios asignados a otras obras o materiales análogos si los hubiere y cuando no, se discutirán entre el Ingeniero Director y el contratista, sometiéndolos a la aprobación de la Dirección. Los nuevos precios convenidos por uno u otro procedimiento, se sujetarán siempre al establecido en el punto anterior.

11. Cuando el contratista, con autorización del Ingeniero Director de obras, emplee materiales de calidad más elevada o de mayores dimensiones de lo estipulado en el proyecto, o sustituya una clase de fabricación por otra que tenga asignado mayor precio o ejecute con mayores dimensiones cualquier otra parte de las obras, o en general, introduzca en ellas cualquier modificación que sea beneficiosa a juicio del Ingeniero Director de obras, no tendrá derecho sin embargo, sino a lo que le correspondería si hubiera realizado la obra con estricta sujeción a lo proyectado y contratado.

12. Las cantidades calculadas para obras accesorias, aunque figuren por partida alzada en el presupuesto final (general), no serán abonadas sino a los precios de la contrata, según las condiciones de la misma y los proyectos particulares que para ellas se formen, o en su defecto, por lo que resulte de su medición final.

13. El contratista queda obligado a abonar al Ingeniero autor del proyecto y director de obras así como a los Ingenieros Técnicos, el importe de sus respectivos honorarios

facultativos por formación del proyecto, dirección técnica y administración en su caso, con arreglo a las tarifas y honorarios vigentes.

14. Concluida la ejecución de la obra, será reconocida por el Ingeniero Director que a tal efecto designe la empresa.

15. La garantía definitiva será del 4 del presupuesto y la provisional del 2 %.

16. La forma de pago será por certificaciones mensuales de la obra ejecutada, de acuerdo con los precios del presupuesto, deducida la baja si la hubiera.

17. La fecha de comienzo de las obras será a partir de los 15 días naturales del replanteo oficial de las mismas y la definitiva, al año de haber ejecutado la provisional, procediéndose si no existe reclamación alguna, a la reclamación de la fianza.

18. Si el contratista al efectuar el replanteo, observase algún error en el proyecto, deberá comunicarlo en el plazo de quince días al Ingeniero Director de obras, pues transcurrido ese plazo será responsable de la exactitud del proyecto.

19. El contratista está obligado a designar una persona responsable que se entenderá con el Ingeniero Director de obras, o con el delegado que éste designe, para todo relacionado con ella. Al ser el Ingeniero Director de obras el que interpreta el proyecto, el contratista deberá consultarle cualquier duda que surja en su realización.

20. Durante la realización de la obra, se girarán visitas de inspección por personal facultativo de la empresa cliente, para hacer las comprobaciones que se crean oportunas. Es obligación del contratista, la conservación de la obra ya ejecutada hasta la recepción de la misma, por lo que el deterioro parcial o total de ella, aunque sea por agentes atmosféricos u otras causas, deberá ser reparado o reconstruido por su cuenta.

21. El contratista, deberá realizar la obra en el plazo mencionado a partir de la fecha del contrato, incurriendo en multa, por retraso de la ejecución siempre que éste no sea debido a causas de fuerza mayor. A la terminación de la obra, se hará una recepción provisional previo reconocimiento y examen por la dirección técnica, el depositario de efectos, el interventor y el jefe de servicio o un representante, estampando su conformidad el contratista.

22. Hecha la recepción provisional, se certificará al contratista el resto de la obra, reservándose la administración el importe de los gastos de conservación de la misma hasta su recepción definitiva y la fianza durante el tiempo señalado como plazo de garantía. La recepción definitiva se hará en las mismas condiciones que la provisional, extendiéndose el acta correspondiente. El Director Técnico propondrá a la Junta Económica la devolución de la fianza al contratista de acuerdo con las condiciones económicas legales establecidas.

23. Las tarifas para la determinación de honorarios, reguladas por orden de la Presidencia del Gobierno el 19 de Octubre de 1961, se aplicarán sobre el denominado en la actualidad "Presupuesto de Ejecución de Contrata" y anteriormente llamado "Presupuesto de Ejecución Material" que hoy designa otro concepto.

Condiciones particulares

La empresa consultora, que ha desarrollado el presente proyecto, lo entregará a la empresa cliente bajo las condiciones generales ya formuladas, debiendo añadirse las siguientes condiciones particulares:

1. La propiedad intelectual de los procesos descritos y analizados en el presente trabajo, pertenece por entero a la empresa consultora representada por el Ingeniero Director del Proyecto.

2. La empresa consultora se reserva el derecho a la utilización total o parcial de los resultados de la investigación realizada para desarrollar el siguiente proyecto, bien para su publicación o bien para su uso en trabajos o proyectos posteriores, para la misma empresa cliente o para otra.

3. Cualquier tipo de reproducción aparte de las reseñadas en las condiciones generales, bien sea para uso particular de la empresa cliente, o para cualquier otra aplicación, contará con autorización expresa y por escrito del Ingeniero Director del Proyecto, que actuará en representación de la empresa consultora.

4. En la autorización se ha de hacer constar la aplicación a que se destinan sus reproducciones así como su cantidad.

5. En todas las reproducciones se indicará su procedencia, explicitando el nombre del proyecto, nombre del Ingeniero Director y de la empresa consultora.

6. Si el proyecto pasa la etapa de desarrollo, cualquier modificación que se realice sobre él, deberá ser notificada al Ingeniero Director del Proyecto y a criterio de éste, la empresa consultora decidirá aceptar o no la modificación propuesta.

7. Si la modificación se acepta, la empresa consultora se hará responsable al mismo nivel que el proyecto inicial del que resulta el añadirla.

8. Si la modificación no es aceptada, por el contrario, la empresa consultora declinará toda responsabilidad que se derive de la aplicación o influencia de la misma.

9. Si la empresa cliente decide desarrollar industrialmente uno o varios productos en los que resulte parcial o totalmente aplicable el estudio de este proyecto, deberá comunicarlo a la empresa consultora.

10. La empresa consultora no se responsabiliza de los efectos laterales que se puedan producir en el momento en que se utilice la herramienta objeto del presente proyecto para la realización de otras aplicaciones.

11. La empresa consultora tendrá prioridad respecto a otras en la elaboración de los proyectos auxiliares que fuese necesario desarrollar para dicha aplicación industrial, siempre que no haga explícita renuncia a este hecho. En este caso, deberá autorizar expresamente los proyectos presentados por otros.

12. El Ingeniero Director del presente proyecto, será el responsable de la dirección de la aplicación industrial siempre que la empresa consultora lo estime oportuno. En caso contrario, la persona designada deberá contar con la autorización del mismo, quien delegará en él las responsabilidades que ostente.