Expert Systems with Applications 42 (2015) 7549–7559

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Data mining for fuzzy diagnosis systems in LTE networks

Emil J. Khatib [a], Raquel Barco [a,*], Ana Gómez-Andrades [a], Pablo Muñoz [a], Inmaculada Serrano [b]

[a] Universidad de Málaga, Communications Engineering Dept., Málaga, Spain
[b] Ericsson, PBO RA Continuous Analysis, Spain

ARTICLE INFO

Article history:
Available online 1 June 2015

Keywords:
Self-healing
Self-Organizing Networks
LTE
Data mining
Data driven learning
Supervised learning
Fault management
Fuzzy systems
Big Data

ABSTRACT

The recent developments in cellular networks, along with the increase in services, users and the demand of high quality have raised the Operational Expenditure (OPEX). Self-Organizing Networks (SON) are the solution to reduce these costs. Within SON, self-healing is the functionality that aims to automatically solve problems in the radio access network, at the same time reducing the downtime and the impact on the user experience. Self-healing comprises four main functions: fault detection, root cause analysis, fault compensation and recovery. To perform the root cause analysis (also known as diagnosis), Knowledge-Based Systems (KBS) are commonly used, such as fuzzy logic. In this paper, a novel method for extracting the Knowledge Base for a KBS from solved troubleshooting cases is proposed. This method is based on data mining techniques as opposed to the manual techniques currently used. The data mining problem of extracting knowledge out of LTE troubleshooting information can be considered a Big Data problem. Therefore, the proposed method has been designed so it can be easily scaled up to process a large volume of data with relatively low resources, as opposed to other existing algorithms. Tests show the feasibility and good results obtained by the diagnosis system created by the proposed methodology in LTE networks.

1. Introduction

In recent years, mobile communications have grown both in traffic volume and offered services. This has increased the expectations for quality of service. In this scenario, service providers are pressed to increase their competitiveness, and therefore to increase quality and reduce costs in the maintenance of their networks. This task can only be achieved by increasing the degree of automation of the network. The Next Generation Mobile Networks (NGMN) Alliance (NGMN, 2006) defined *Self-Organizing Networks (SON)* as a set of principles and concepts to add automation to mobile networks. Recently, these ideas were applied to *Long Term Evolution (LTE)* networks by the 3rd Generation Partnership Project (3GPP) (3GPP, 2012) in the form of use cases and specific SON functionalities. The SON concept is composed by three fields: *self-configuration*, *self-optimization* and *self-healing*. This paper is focused on self-healing, which includes all the functionality targeted towards automating troubleshooting in the radio access network (RAN). Currently, the task of RAN troubleshooting is manually done, so the ability of automating it is a great competitive advantage for operators. The benefits of self-healing are numerous, since it will offload troubleshooting experts of repetitive maintenance work and let them focus on improving the network. It also reduces the downtime of the network, therefore increasing the quality perceived by the users.

Self-healing is composed of four main tasks: fault detection, root cause analysis (diagnosis), compensation and fault recovery (Barco, Lázaro, & Muñoz, 2012). Currently there are no commercial implementations of these functions, although some research effort has been recently made.

The reason of this shortage of implementations, according to the COMMUNE project (COMMUNE, 2012) is the high degree of uncertainty in diagnosis in the RAN of mobile networks. The COMMUNE project uses a case based reasoning algorithm (Szilagyi & Novaczki, 2012) where a vector of Performance Indicators (PIs) is compared against a database of known problems. The cause will be the same as the case of the nearest stored problem.

In Barco, Díez, Wille, and Lázaro (2009) a Bayesian Network (BN) is used to do the diagnosis. To implement the system, it is required that an expert sets the parameters of the BN, so a Knowledge Acquisition Tool is also presented in Barco, Lázaro, Wille, Díez, and Patel (2009).

* Corresponding author.
E-mail addresses: emil@uma.es (E.J. Khatib), rbarco@uma.es (R. Barco), aga@ic.uma.es (A. Gómez-Andrades), pabloml@ic.uma.es (P. Muñoz), inmaculada.serrano@ericsson.com (I. Serrano).

http://dx.doi.org/10.1016/j.eswa.2015.05.031
0957-4174/

The UniverSelf project (Univerself, 2012) combines BNs with case-based reasoning (Bennacer, Ciavaglia, Chibani, Amirat, & Mellouk, 2012; Hounkonnou & Fabre, 2012) for the diagnosis process. Apart from the previous references in diagnosis, research in self-healing has also been extended to detection (Asghar, Hamalainen, & Ristaniemi, 2012; Barreto, Mota, Souza, Frota, & Aguayo, 2005) and compensation (Eckhardt et al., 2011; Razavi, 2012), which are out of the scope of this paper.

This paper proposes a method for learning troubleshooting rules for diagnosis methods based on Fuzzy Logic Controllers (FLCs) (Lee, 1990). FLCs use fuzzy logic (Zadeh, 1965) to assign fuzzy values to numerical variables, and by applying fuzzy rules, they obtain the value of an output variable (e.g. a parameter value) or a given action. FLCs have been used for diagnosis in other fields, such as industrial processes (Serdio, Lughofer, Pichler, Buchegger, & Efendic, 2014), machinery operation (Serdio et al., 2014; Lughofer & Guardiola, 2008) or medical diagnosis(Innocent, John, & Garibaldi, 2004). Although FLCs have been used in mobile networks for self-optimization (Muñoz, Barco, & de la Bandera, 2013a; Muñoz, Barco, & de la Bandera, 2013b), there are no previous references proposing its application to self-healing.

The implementation of the diagnosis process is done by Knowledge-Based Systems (KBS) (Akerkar & Sajja, 2010; Triantaphyllou & Felici, 2006) such as BNs and FLCs. KBS are composed of two main parts: the Inference Engine (IE) and the Knowledge Base (KB). This separation permits the algorithm to be used in a variety of situations by changing only the KB. Still, the construction of the KB, or Knowledge Acquisition (KA) (Triantaphyllou & Felici, 2006; Maier, 2007) is a major research topic. This is where the previous KA proposals in literature for diagnosis in wireless networks fail to deliver convincing results. The KA process involves the troubleshooting experts in a time consuming process (Studer, Benjamins, & Fensel, 1998; Ruiz-Aviles, Luna-Ramírez, Toril, & Ruiz, 2012; Chung, Chang, & Wang, 2012; Barco et al., 2009). It is based on the fact that the knowledge is contained in the experience of the expert. An alternative approach (Triantaphyllou & Felici, 2006) considers that the knowledge applied by the experts in the troubleshooting process will also be contained in its results. Every pair composed of the PIs and the fault cause and/or action(s) to be taken provided by the expert contains information about what aspects are observed by the expert and how they are related to each other. Therefore, a problem database (Hatamura, Ilno, Tsuchlya, & Hamaguchi, 2003) can be created, where each problem is saved along with its diagnosis; and this database will hold the expert's knowledge implicitly.

Data mining (DM) consists of the discovery of patterns in large data sets through the application of specific algorithms (Kantardzic, 2011; Witten, Frank, & Mark, 2011; Han, Kamber, & Pei, 2012). The result of DM is a model of the studied system. DM is used to process information from sensor networks (Papadimitriou, Brockwell, & Faloutsos, 2003), in scientific data collection (Ball & Brunner, 2010) or computer science (Ektefa, Memar, Sidi, & Affendey, 2010). DM is also used in commercial applications to find marketing trends (Linoff & Berry, 2011). Modern data collection and monitoring systems produce large amounts of data containing valuable knowledge, but due to the huge amount of information, this knowledge is hidden and needs to be extracted and easily visualized. In fact, the traditional DM techniques are often insufficient or ineffective for the large amount of available data. This leads to the development of Big Data techniques (Russom, 2011), which deal with databases that have special requirements due to one or more of the following factors, known as the *3 Vs of Big Data*:

- Volume: the amount of data generated may require new transmission, storage and processing techniques. In the case of mobile networks, data is produced by each network element (e.g. eNodeB) over all the network, but also by each connected terminal.
- Velocity: usually there are constraints on the time when the analysis results must be available, so that the information can be used in real time. In mobile networks, part of the data is produced in streams, that is, continuously as new events happen while the users access the services. In order to reduce processing times, heavy parallelization is very often the best solution.
- Variety: data is not always normalized and clearly structured. Also, formats vary depending on the nature of the data and the element that produces it.

In this work, a DM algorithm for obtaining fuzzy rules in the "*if* ... *then* ..." form for the diagnosis of the RAN of LTE networks is proposed, which relates certain behaviors in the PIs of a sector at a given time to the possible problem that is present in it. The rules reflect the knowledge of the experts that is implicitly contained in the results of their work (a database containing the solved problems). Other learning algorithms have been proposed for fuzzy rules in other fields, but none of them have been adapted to mobile communication networks (Wang & Mendel, 1992; Cordon, Herrera, & Villar, 2011; Lughofer & Kindermann, 2010; Pratama, Anavatti, Angelov, & Lughofer, 2014; Chen, 2013; Babuska, 1998). This algorithm has been designed to be easily parallelizable in order to fit the velocity requirements described earlier and to minimize the memory footprint so a large volume of data can be processed.

The remaining of this paper is organized as follows. In Section 2 the traditional processes of troubleshooting is presented, along with the use of KBS for automating it. In Section 3 the proposed DM algorithm for automating the KA process is presented. Afterwards, in Section 4, the system is tested and its performance is compared with another learning algorithm. Finally, the conclusions are discussed in Section 5.

2. Problem formulation

2.1. Troubleshooting in LTE networks

The process of troubleshooting is made up of four main tasks:

- Detection: determining that there is a problem in the network and isolating its origin, that is, the sectors that are degrading the performance of the whole network.
- Compensation: reconfiguration of the neighboring sectors to cover the affected users.
- Diagnosis: determining the cause of the problem.
- Recovery: actions to restore the affected sector to a normal state.

This process is usually manually done. Experts monitor a reduced set of Key Performance Indicators (KPIs), aggregated over the whole network. KPIs are composite variables that contain the information of several low-level PIs and reflect the general behavior of the sector. When one (or several) of those KPIs is degraded (e.g. it is lower than a certain threshold), a list of the worst offenders is obtained, showing the sectors that more strongly degrade the KPI. Those sectors are then diagnosed by observing further symptoms (i.e. abnormal values of PIs and other low-level metrics, such as counters, measurements and alarms). The relations among these symptoms, described by heuristic "*if* ... *then* ..." rules, determine