Expert Systems With Applications 64 (2016) 56–68

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Combination of multiple diagnosis systems in Self-Healing networks

David Palacios, Emil J. Khatib, Raquel Barco*

Communications Engineering Dept., University of Málaga, 29071, Málaga, Spain

ARTICLE INFO

Article history:
Received 8 January 2016
Revised 20 July 2016
Accepted 21 July 2016
Available online 22 July 2016

Keywords:
LTE
Self-healing
Root cause analysis
Self-organizing networks (SON)
Hybrid ensemble classifier
Automatic fault identification

ABSTRACT

The Self-Organizing Networks (SON) paradigm proposes a set of functions to automate network management in mobile communication networks. Within SON, the purpose of Self-Healing is to detect cells with service degradation, diagnose the fault cause that affects them, rapidly compensate the problem with the support of neighboring cells and repair the network by performing some recovery actions.

The diagnosis phase can be designed as a classifier. In this context, hybrid ensembles of classifiers enhance the diagnosis performance of expert systems of different kinds by combining their outputs. In this paper, a novel scheme of hybrid ensemble of classifiers is proposed as a two-step procedure: a modeling stage of the baseline classifiers and an application stage, when the combination of partial diagnoses is actually performed. The use of statistical models of the baseline classifiers allows an immediate ensemble diagnosis without running and querying them individually, thus resulting in a very low computational cost in the execution stage.

Results show that the performance of the proposed method compared to its standalone components is significantly better in terms of diagnosis error rate, using both simulated data and cases from a live LTE network. Furthermore, this method relies on concepts which are not linked to a particular mobile communication technology, allowing it to be applied either on well established cellular networks, like UMTS, or on recent and forthcoming technologies, like LTE-A and 5G.

1. Introduction

The growing demand for mobile services with ever-increasing bandwidth and the expanding number of users make necessary the deployment of new and more efficient mobile communication networks over the existing ones (GSM, UMTS), such as Long-Term Evolution (LTE). However, the complexity of this heterogeneous scenario, which comprises several Radio Access Technologies (RAT), requires challenging maintenance and complex operational tasks. Mobile operators need to offer new demanding services without increasing either operational expenditures (OPEX) or capital expenditures (CAPEX). In order to deal with that problem, the 3rd Generation Partnership Project (3GPP) has proposed Self-Organizing Networks (SON) (3GPP (d)) as networks that include mechanisms to automate network procedures in order to help mobile operators with their management work, providing significant cost reduction. This automation of network management will also be essential in near and future technologies, like LTE-Advanced and 5G (3GPP (b)).

SON comprises three groups of functions: Self-Configuration, Self-Optimization and Self-Healing. The aim of the latter is to autonomously solve the problems that a cell, with service degradation or outage, could present (3GPP (e); Barco, Lázaro, and Muñoz (2012)). This is done by means of four stages:

- *Fault Detection:* Responsible for finding cells with problems, i.e., cells experiencing service outage or just suffering an unacceptable service degradation.
- *Diagnosis of the fault cause:* In this step, the actions to be performed in order to recover the system from the degradation it is suffering are decided. This step can be divided into two substages: Fault Identification, this is, identifying the fault cause based on observable symptoms such as Key Performance Indicators (KPI) and alarms; and Action Identification, which corresponds to the decision of what tasks to perform to recover the system normal performance.
- *Fault recovery:* In this step, the proposed solutions are carried out.

* Corresponding author. Fax: +34952132027.
E-mail addresses: dpc@ic.uma.es (D. Palacios), emil@uma.es (E.J. Khatib), rbarco@uma.es (R. Barco).

http://dx.doi.org/10.1016/j.eswa.2016.07.030

- *Fault compensation:* Since diagnosing the fault and repairing it normally takes some time, compensation aims to diminish the impact of the fault by changing parameters in neighboring cells.

This paper is focused on the diagnosis task, in particular in the fault identification, also called root cause analysis. Once a problem has been detected in a cell, root cause analysis identifies the fault cause given the value of performance indicators, alarms, counters, mobile traces, etc. In the context of cellular networks, some diagnosis systems have been recently proposed. Barco, Díez, Wille, and Lázaro (2009) and Barco, Lázaro, Wille, Díez, and Patel (2009) proposed diagnosis systems based on Bayesian Networks. Szilágyi and Nováczki (2012) used a scoring system in order to determine how well a specific case fits a diagnosis. Nováczki (2013) enhanced the previous system by adding profiling techniques. The method in Khatib, Barco, Gómez-Andrades, and Serrano (2015) was based on fuzzy logic and genetic algorithms. Gómez-Andrades, Muñoz, Serrano, and Barco (2016) proposed a diagnosis system based on Self-Organized Maps (SOM).

Each of the previous methods has its pros and its cons. In practice, this makes the selection of the diagnosis technique cumbersome when the aim is to deploy a automatic diagnosis system in a real network. Furthermore, once the technique has been decided, e.g., fuzzy logic, operators normally design several standalone diagnosis models. This is due to the fact that, firstly, different troubleshooting experts will build different models and secondly, when models are learnt from historical cases, different training datasets will result in different models.

To cope with the limitations of standard classifying systems in terms of accuracy and dataset-dependent performance, ensembles of classifiers arose. Within these, homogeneous and heterogeneous (commonly known as hybrid) ensembles of classifiers may be found, where the former stand for the ensemble of classifiers of the same kind and the latter stand for the combination of different kinds of systems and datasets. Despite homogeneous ensembles have been widely studied and as of today still are extensively used in different fields (Begum, Chakraborty, & Sarkar, 2015; Liu, Chen, Song, & Han, 2009; Shen & Chou, 2006; Wiezbicki & Ribeiro, 2016). In this paper, a method for the generalized combination of multiple diagnosis systems based on a hybrid ensemble approach is proposed and tested in the context of cellular networks, which to the authors' knowledge is a research area still to be explored. The proposed work describes a method to gather, combine and use the knowledge held by any kind of expert system in any field that makes use of a classifying or diagnosis system. In this work, the proposed method is applied in the fault cause diagnosis in cellular networks, where the expertise may be provided either by a human troubleshooting expert or by a database of cases assessed by automatic diagnosis systems. The proposed method allows combining diagnosis systems in a wide sense, being able to merge both several diagnosis models (expertise) and the tools used for their application (automatic diagnosis techniques) in the form of supervised or unsupervised classifying systems.

Up to now, hybrid ensembles of classifiers are mainly based on a set of baseline systems which must first assess the cases under test and, consequently, provide partial diagnoses which are finally combined into a final decision using a majority vote scheme (Ciocarlie, Lindqvist, Nováczki, & Sanneck, 2013; Gandhi & Pandey, 2015; Wei et al., 2014). This procedure requires a relatively high number of diagnosis techniques to be run in the test stage and, therefore, a noticeable expenditure of computational and time resources. The proposed work, however, presents a method which allows combining the diagnoses that the standalone diagnosis systems would output for a case under test without actually needing them to be run, thus lightening the computational weight of the test stage.

The main contributions of this paper are:

- A method to combine any number and kind of different standalone classifiers as well as different sources of expert knowledge in order to get an enhanced performance compared to that of the base classifiers. In the context of troubleshooting in cellular networks this comprises the combination of several diagnosis models and techniques for the automatic diagnosis.
- A method to lighten the computational cost of the evaluation stage in hybrid ensembles of classifiers. This work proposes a scheme to model and emulate the behavior of every standalone classifier so these need not to be continuously queried before combining their partial diagnoses.

This paper is organized as follows. Section 2 presents the problem formulation. Section 3 introduces the proposed method for combining multiple baseline diagnosis systems. In Section 4 results are analyzed by means of both a network simulator and data from a live LTE network. In Section 5 the future lines of work are outlined. Finally, Section 6 summarizes the main conclusions.

2. Problem formulation

2.1. Root cause analysis in mobile communications networks

In the same way that a patient is diagnosed by a doctor based on the symptoms he shows, the status of a communications network may be diagnosed based on a set of performance indicators. This diagnosis task, also called root cause analysis or troubleshooting, is often carried out by human experts using their knowledge on the underlying relations that the observed indicators and the status of the network have. However, the number of symptoms (counters, alarms, KPIs, call traces, etc.) and possible fault causes the expert has to deal with increases as networks grow in size and complexity, which makes this task to become a very difficult and time consuming issue.

Furthermore, the current manual troubleshooting is a layered task, guided by a Trouble Ticket (TT) system. In this problem solving system, a group of specialists tries first to diagnose and solve the problem by performing some simple checks. If they can not find the root of the problem, this is raised to a more specialized team (and so on), which performs a deeper study on the symptoms the case exhibits and resorts to field engineers in case they need to make some on site checks.

As a response to this more and more inefficient procedure, automatic diagnosis systems arose in an attempt of imitating the way of acting of troubleshooters. Fig. 1 shows the basic scheme of a system for automatic diagnosis. It is composed of an automatic diagnosis technique and a diagnosis model. The first is an artificial intelligence system that outputs a diagnosis taking a set of symptoms, e.g., (KPIs) from a test case as its input. The second represents the knowledge a human expert would have on the underlying relations between the symptoms and the fault causes and may take different forms depending on the diagnosis technique it is destined to work with. For example, a diagnosis model may consist of the parameters (e.g., prior probabilities and probability density functions) required by a given diagnosis technique (e.g., bayesian classifier) or a set of rules for other techniques (e.g., Case Base Reasoning, CBR). As it can be seen in this figure, the diagnosis model may be built from a set of training cases by means of a machine learning algorithm or by troubleshooting experts by gathering their knowledge. The proposed method aims to combine the knowledge acquired by any number and kind of diagnosis models and automatic diagnosis techniques in an attempt to reduce the errors in fault detection and diagnosis.