Knowledge-Based Systems 25 (2012) 77–87

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Interpretable knowledge extraction from emergency call data based on fuzzy unsupervised decision tree

Francisco Barrientos [a,*], Gregorio Sainz [a,b]

[a] CARTIF Centro Tecnológico, Parque Tecnológico de Boecillo 205, 47151 Valladolid, Spain
[b] University of Valladolid, Systems Engineering and Control Department, School of Industrial Engineering, 47011 Valladolid, Spain

ARTICLE INFO

Article history:
Available online 1 March 2011

Keywords:
Unsupervised decision trees
Fuzzy clustering
Fuzzy rules
Interpretability
Emergency call center
Decision support systems

ABSTRACT

Nowadays, call centers are common in different areas of activity providing customer services, medical attention, security services, etc. Each type of call center has its own particularities but all call centers have to plan the availability of resources at each time period to support the incoming calls. The emergency call centers are a special case with extra restrictions.

In this context, this work is devoted to providing support for the decision making about resource planning of an emergency call center in order to reach its mandatory quality of service. This is carried out by the extraction of interpretable knowledge from the activity data collected by an emergency call center.

A linguistic prediction, categorization and description of the days based on the call center activity and information permits the workload for each category of day to be known. This has been generated by a fuzzy version of an unsupervised decision tree (FUDT), merging decision trees and clustering. This involves quality indexes to reach an adequate trade-off between the tree complexity and the category quality in order to guarantee interpretability and performance. This unsupervised approach deals correctly with the real management of this type of centers generating and preserving expert knowledge.

1. Introduction

Nowadays, call centers are a very usual resource in different areas and activities: companies, social services, medical attention, security services, etc. A call center is a service network in which agents provide telephone-based services [1]. The implementation for each case has its own particularities but all call centers must manage their resources to reach efficiency and quality of service. One basic aspect for this goal is to plan the size of the call center and the available resources at each moment to dispatch the incoming calls. In this domain, emergency call centers are a special and particular case of call center, in which harder constraints are introduced to reach better performance in accordance with their delicate nature, the security and quality protocols established and the consequences of an inadequate attention given by the call center.

In this context, the main problem in the design and management of any call center is to achieve an adequate trade-off between quality of service (QoS) and efficiency. A medium-large sized center can manage thousands of calls per day, and each one must be answered in very few seconds [2]. In [3] and [4] reviews about call centers, and their domain, are shown from a management point of view.

In general, most of the approaches to this problem come from the operations management domain. But other approaches are possible, paying attention to the need to approximate the workload for a given period, and then to estimate the availability of resources to respond to this demand. Call centers have usually been modeled by queuing theory, using the paradigm of producer/consumer. In [5] queueing models for multiclass call centers are dealt with.

In [6], the authors illustrate the benefits of the centralization of call centers; to have one emergency call center is better than having one per involved organization (sanitary, police, etc.). An algorithm for capacity sizing a call center is developed in [7] by queueing and stochastics. This method is not valid in this case because the shared resources work in an uncoupled way, i.e. an emergency call center has no decision capability, only information capability. A stochastic programming model for scheduling is introduced in [8]. Time series and soft computing are considered in [9] to forecast inbound calls. In [1] a statistical analysis of a call center is done. In [10], this approach is in focus on taking advantage of knowledge through similar philosophy to our work: the focus is on finding out the relevant attributes in order to reduce the number of the fatality in the aviation incidents.

An emergency call center is designed to efficiently respond to individual emergency calls, assisting call-makers and redirecting

* Corresponding author.
E-mail address: frabar@cartif.es (F. Barrientos).
URL: http://www.cartif.es (F. Barrientos).

0950-7051/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2011.01.014

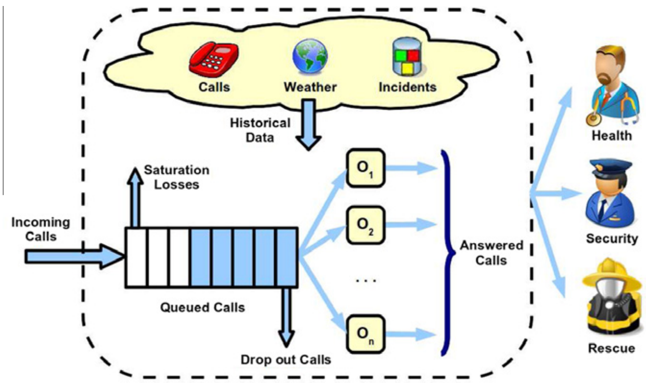

Fig. 1. The goal is to maximize availability, productivity aside, and minimize missed calls and users' waiting time.

their calls to the appropriate emergency service provider (e.g. law enforcement, fire response, or emergency medical services) if necessary [11] (see Fig. 1). The references about emergency call centers and their performance from the point of view of this work are not very usual for evident reasons. In [11] the 9-1-1 call data are used to detect clusters and to correlate with emergency events by space–time permutation scan statistics, but this method is based on the location of the calls, information that is not recorded by the system under study.

In this context, the scope of this work does not involve an exhaustive study and benchmarking of techniques and approaches about emergency call centers. Here, the work is focused on a real and practical point of view: a solution based on the extraction of interpretable knowledge from real data from an emergency call center. The goal to be achieved is a linguistic prediction, categorization and description of each type or category of day, and the workload involved, for the normal working of the emergency call center. Thus, the workload of a day can be anticipated and the resources to be used can be managed.

This solution is based on a fuzzy unsupervised decision tree (FUDT) that has been tuned to generate a set of interpretable fuzzy rules, similar to other domains as in [12], to predict and define clusters or day categories, both from a linguistic point of view, trading-off interpretability, tree complexity and tree performance. This problem is partially involved in [9] by time series but the present work is devoted to a solution to describing day categories, their prediction, their description and the associated workload for each category by linguistic terms based on the emergency call center domain. Unlike in [13], in which the application of an user-feedback based extension of a decision tree induction algorithm is proposed to detect criminals e-mails, and [14], in which the author proposes a new decision trees induction algorithm that reduces the intangible costs when taking into account waiting costs. In our case an unsupervised approach is mandatory in order to deal with weak availability of the expert knowledge.

This paper is organized as follows: Section 2 gives a short description of the proposal of resource management for an emergency call center and surveys the theoretical basis that supports the model. Section 3 develops the proposal and its methodology. The experimentation carried out and the analysis of the results are shown in Section 4. Finally, the main conclusions and further work and research are outlined in Section 5. In Appendix A the rule base and some tables of results are given for a better readability.

2. Emergency call center problem: an approach

This work is focused on emergency calls, paying attention to obtaining the maximum availability of the service in order to avoid missing calls. Here, economic factors are not initially considered. The aim of this approach is to take advantage of the information and data collected from the call center by extraction of knowledge as follows:

- Categories characterizing, and describing, linguistically the different types of days for the normal working of the emergency call center.
- Classification system, based on an interpretable fuzzy rule set, to predict the type or category day, and therefore its workload, as one of the categories described above. In this way, the workload for each day can be estimated in advance, then the resource management to give support needed can be carried out.

Both elements, categories and fuzzy rule set, are generated simultaneously from the collected data in the emergency call data, trading-off amongst different aspects: complexity, interpretability, performance. This interpretable knowledge provides support to decision making about resource management.

In order to carry out this solution, a fuzzy version of the unsupervised decision trees (UDT) [15] is used. Decision trees and clustering are merged in UDT techniques, providing the performance from both points of view that fit perfectly with the aim of this work.

The input data used to characterize the days and to generate the fuzzy rule set are the variables collected by the emergency call center: intrinsic and other extrinsic variables which may be relevant to define the call center workload. In Table 1 a summary of the considered intrinsic an extrinsic variables are shown. These were taken into consideration after brainstorming with emergency call center managers.

"Non_Incidents" is the percentage of daily calls corresponding to void calls, acknowledgements, jokes, etc. *"Information"* are the calls asking for some information. *"Incidents"* shows the percentage of daily calls about real incidents that should be attended by emergency services. This group is subdivided as follows: *"Sanitary"*, sanitary services are required and *"Security"* for police corps. *"Search and Rescue"* involves fire and rescue corps. Finally, *"Basic_Services"* includes other issues like water/gas/electricity supply problems, etc.

Table 1
Input variables summary. Daily statistics.

Num	Name	Type	Min	Max	Mean	Std. dev.
1	Non_Incidents (%)	Intrinsic	64.90	92.70	83.38	4.33
2	Information (%)	Intrinsic	0.30	16.30	1.89	1.35
3	Incidents (%)	Intrinsic	6.00	31.20	14.72	4.34
4	Sanitary (%)	Intrinsic	33.70	75.20	57.03	5.94
5	Security (%)	Intrinsic	18.60	57.10	36.30	5.65
6	Search and Rescue (SAR)(%)	Intrinsic	0.30	24.50	5.04	3.69
7	Basic_Services (%)	Intrinsic	0.00	9.90	1.64	1.35
8	Calls_Midnight	Intrinsic	8.00	495.00	163.33	91.83
9	Calls_Dawn	Intrinsic	0.00	214.00	33.44	21.89
10	Calls_Morning	Intrinsic	92.00	722.00	245.76	98.60
11	Calls_Afternoon	Intrinsic	109.00	835.00	360.07	132.07
12	Calls_Evening	Intrinsic	52.00	770.00	237.51	118.68
13	Total_Calls	Intrinsic	1658.00	9504.00	4788.09	1608.71
14	T_Avg	Extrinsic	−3.37	28.73	11.79	7.68
15	T_Max	Extrinsic	1.30	38.90	19.42	9.01
16	T_Min	Extrinsic	−11.90	17.90	3.92	6.36
17	Rain	Extrinsic	0.00	192.20	6.10	15.07
18	Snow	Extrinsic	0.00	6.00	0.17	0.67
19	Fog	Extrinsic	0.00	6.00	0.54	1.12
20	Day_of_Week	Extrinsic	1.00	7.00	4.00	2.01
21	Month	Extrinsic	1.00	12.00	6.01	3.55