Knowledge Discovery from Trouble Ticketing reports in a large Telecommunication Company

Yaiza Temprado
Telefónica I+D
ytr@moviquity.com

Carolina García
Telefónica I+D
cgv@tid.es

Francisco Javier Molinero
Telefónica I+D
fjmv@tid.es

Julia Gómez
Telefónica I+D
juliag@tid.es

Abstract

This paper describes the work developed by Telefónica I+D about an application of advanced Data Mining, Text Mining and Machine Learning techniques for the study of the network elements failures managed by the Trouble Ticketing System of a large telecommunication company, in order to be able to analyze, prioritize and, in some cases, solve without human intervention the huge amount of trouble reports to be managed. Furthermore, this paper will present the techniques used for its achievement, as well as the results obtained so far, showing how these techniques may help important companies to save plenty of time and resources in fault management, improving the service quality.

1. Introduction

Traditionally, important companies have jealously stored their business data, using them only for reporting and auditing. However, in recent years companies have started to realize that the huge amount of available information offers a much broader scope. That is the reason why Data Mining techniques are quickly expanding, providing companies with key information. The present research wants to take a further step, with the aim not only to get relevant information, but also to use this information in order to improve and optimize the Business Processes of a Telecommunication Company.

In our case, we will focus on the Maintenance Process and will use the data of Trouble Ticket reports of the telecommunication network (i.e. tickets that help to manage all the problems, failures and breakdowns that daily happen in the network). Our goal is to optimize the solving process, shortening the necessary time to do so and automating some of the tasks needed. Besides, our two main objectives are: Increasing the client's satisfaction, since failures are solved quicker and reducing costs, since available resources can be optimized.

In large companies' infrastructures, hundreds of thousands of trouble tickets can be generated every year. These tickets can be a consequence of different factors: faulty hardware or software, breakdowns, unusual processes perceived by the supervisors or customers complaining about the service.

Due to the great number of problems, these companies have been using complex systems over the years in order to automate tasks and facilitate their management, saving resources and costs. These systems are known as Trouble Ticketing Systems (or TT-Systems). Nevertheless, lots of people and time are still required to process the tickets and solve the problems. Some of the tickets can be created and managed by automatic supervision systems, but others, due to their complexity, require specialized technicians for their creation. This implies two kinds of sources for the ticket creation: the automatic one and the manual one. The later provides more specific data about the failure source, but it also means that tickets are described using a natural language, with an inherent ambiguity.

On the other hand, most of the problems can be solved remotely nowadays. However, some of them still require work on-site, meaning that a specialized technician must travel to fix the problem, which obviously increases the final cost. Consequently, it is essential to avoid wrong diagnoses, preventing works on-site when it is not necessary. Predicting the need of works on-site in order to optimize planning, has been

978-0-7695-3514-2/08 $25.00 © 2008 IEEE
DOI 10.1109/CIMCA.2008.116

one of the goals of the work developed by Telefónica I+D and will be further discussed in this paper.

Moreover, due to the high volume of tickets that must be daily dealt with, and the limited number of resources, it is necessary to provide a way to prioritize tickets, so that supervisors can decide what tickets are more critical and should be solved sooner. But this is not the only fact to take into account, since a ticket first qualified as unimportant, could eventually become into a dangerous one. So, another key goal of Telefónica I+D's work has been to predict future increases in the ticket priority, allowing supervisors to improve planning and detect possible problems.

2. Background

In companies with a huge network infrastructure, optimizing the response given whenever a failure is encountered is critical. Over the years, many telecommunication companies, as well as companies from other different sectors, have tried to create systems capable of recognizing failure symptoms in order to speed up and automate ticket creation, advising system users when taking decisions, etc.

Until mid-90s, Expert Systems were the prevailing technique to address this type of problems. [1], [2], and [3] are just some examples of this type of approach, that although laid the foundations of the work that has been done in fault management hitherto, presented important limitations, such as the lack of flexibility for considering new cases or the need of an expert to encode the knowledge into the system. The lack of experts, together with the inflexibility inherent to this solution, does not render this approach as the most appropriate one in a scenario where networks are in constant evolution.

In recent years, more advanced techniques coming from Machine Learning are being used, such as Neural Networks, Decision Trees, Bayesian Networks or Genetic Algorithms. These techniques solve the aforementioned problems of Expert Systems, being able to obtain knowledge automatically (without the intervention of an expert) and, when properly used, react very well when treating new cases or noisy data [4][5].

Although there are known cases where companies have applied Machine Learning techniques for the automatic creation of tickets (for example, searching for unusual behaviour patterns in the network traffic [6][7]), there are not known cases where these techniques have been applied in a later step, once the ticket has been created. I.e. whereas many companies have applied Machine Learning before the process known as Trouble Ticketing (ticket management workflow), the present work addresses the Trouble Ticketing process itself, shortening time and optimizing resources in ticket resolution, and also improving customer satisfaction.

3. Data preprocessing and feature selection

The initial available data at the beginning of the project consist of already solved tickets (closed), which are registered in the historical repository of the Trouble Ticketing System. The information the repository provides from each ticket refers to the moment when it was closed, as well as all the different actions performed from its creation to its closing.

These actions refer to the creation of the ticket, its transference to other involved departments, its association with another ticket (so that the solution of one of them induces the solution of the other), its closing, etc. Therefore, and due to the dynamic management or the ticket workflow, it is clear that there are some data which may change since the creation moment.

Given that the aim of this project is predicting different parameters during the life cycle of an active ticket, whereas for the execution of our experiments only closed tickets are available, the learning process of our system requires a first step to reconstruct the evolution of each ticket data so that it can later be used for prediction.

In order to achieve this, we start with the ticket in its initial situation or creation time (instant t_0). This starting point can be reconstructed because there is information in the repository that is known to remain invariable from creation to closing. From this point, by means of the information about the actions also stored in the repository (that includes dates, who did them, etc.), the different snapshots of the ticket can be reconstructed up to the closing action.

$$\text{Ticket}\ (t_0, \text{action}_0) \rightarrow \text{Ticket}\ (t_1, \text{action}_1) \rightarrow \ldots \rightarrow \text{Ticket}\ (t_n, \text{action}_n)$$

Instants t_i match the date and time of the actions obtained from the repository; this fact makes it possible to work on the prediction of information in a given instant of time as if it were information recovered in real time.

The attributes extracted from a ticket can be classified in three types: attributes describing the equipment or element affected by the breakdown, attributes describing the breakdown itself and attributes containing dynamic information that reflects the situation of the breakdown at each moment.

Of all the available information about a ticket, only those attributes directly related to the symptom contain information about the possible cause of the problem. Unfortunately, this information is filled in by technicians who perform the same operations many times a day, tending to do it quickly and without much detail, so that these attributes do not always contain complete or relevant information. Besides, these attributes are filled in with natural text, which is not directly manageable by machine learning techniques. Consequently, they need to be carefully analysed in order to extract as much information as possible. Some Text Mining techniques have been used, such as stemmer algorithms, entity-relationship models, frequency recount, or stop lists [8].

In order to reduce the huge number of attributes obtained, it has been necessary to perform an attribute selection to keep the most relevant ones. To that end, wrappers [9] have been used in combination with those machine learning algorithms that are more suitable for the solution of these particular problems. Wrappers search for a subset of features guided by the behaviour of a concrete classifier, which is the best option at this step of the problem resolution, since wrappers have demonstrated to obtain reasonably better results than more traditional statistical methods for feature selection (e.g., χ-square) [10].

Of course, once the selected algorithms for feature selection have been applied, the expert knowledge is also used, in order to polish the attributes values and add new features. Generally, the expert knowledge helps to increase the data quality the system is going to work with.

4. The problem of predicting the need of work on-site.

Due to the network size and the huge number of customers, several thousands of failures occur and, therefore, several thousands of tickets are daily created in the TT System. In this context, any automation or help means enormous savings in time and money, and the corresponding improvement of service quality. During a ticket life cycle, the person responsible for its resolution can carry out different actions or tasks. One of the main tasks is deciding whether a ticket can be solved remotely or a technician must travel to fix the failure on-site. So, two problems must be addressed:

- When the ticket is created, the system must try to predict whether the ticket will require working on-site at some moment in its life cycle.
- During the ticket life cycle, the system must try to predict whether the next action to be executed in its workflow is sending a technician to work on-site.

Naturally, both problems are perfectly compatible: on the one hand, when the ticket is created an initial probability that the ticket requires working on-site will be given. As the ticket workflow evolves and the information is being completed, there will be more accuracy in predicting the moment when this action is needed.

4.1. Training and Classification

To solve this particular problem, the tickets are classified according to the type of network (transmission, radio …) because the failures can be completely different. From data analysis it can be easily observed that attribute information varies depending on the creation type (manual or automatic). Thus, both cases are treated separately.

Something that must be emphasized is the fact that the increase of remote control systems reduces the number of tickets that require working on-site, with values around 10%. So, results must be carefully analyzed. Sending a technician means an important added cost for the company, so it is essential not to send technicians whenever works can be done remotely, i.e. in this particular case, false positives are very expensive and, therefore, must be penalized more than false negatives. Bearing this in mind, the classification algorithms have been combined with a cost matrix, allowing a penalization adjustment.

For the resolution of this problem different algorithm families have been tested: statistical, bayes-based, trees, and rules. The table below shows the best results obtained with the algorithms selected from each family:

	Tickets created automatically		Tickets created manually	
	Acc.(%)	MAE	Acc.(%)	MAR
Decision table	93.2	0.112	93.6	0.173
Tree C4.5	94.1	0.091	93.4	0.095
Bayes AODE	92.1	0.088	93	0.079

Figure 1. Results for work on-site prediction: accuracy (Acc.) and mean absolute error (MAE)

But not only Precision and Recall must be considered when selecting the algorithm that best fits our needs; as mentioned above, false positives are more expensive than false negatives, so that this fact must be a determinant factor when selecting the algorithm.

	Tickets created automatically		Tickets created manually	
	%FP	%FN	%FP	%FN
Decision table	2.4	67.1	1.4	48.8
Tree C4.5	1.4	63.9	1.7	48.5
Bayes AODE	2.8	74.3	2.6	44.6

Figure 2. False positives (FP) and negatives (FN)

In view of the results, the algorithm that offers the best results is C4.5, not only because of its better accuracy and smaller mean error, but also because of its noise resistance and the possibility of tracing the causes of each classification, in order to improve the process and detect possible errors.

5. The problem of predicting severity escalation.

When a ticket is created, be it automatically by a supervision agent, or manually by specialized technicians, it is assigned an initial severity; i.e. from the information about the problem that is the cause of the ticket (like the number of affected customers or the type of network) it is assigned a value between 0 and 9 indicating its priority, as well as a degree of importance.

This initial severity is not fixed: sometimes, at the moment of the ticket creation, some data related to the scope of the problem are missing or incomplete, so that the ticket might be more important than initially expected. This change in the initial severity is called escalation. An escalation can be caused by two reasons: either the initial estimation was not correct, or having a correct initial low severity, and thus not urging its resolution, it remains open (not solved) longer than the standard time for similar tickets, so that it escalates to a higher severity.

This second case is where Machine Learning techniques can play a more important role because if it is possible to predict when a low relevance ticket is going to escalate and become a high severity ticket, supervisors can be warned of this situation in advance. With this goal in mind, this complex problem has been divided into four sub-problems:

- Predicting, at the time of a ticket creation, whether the ticket is going to suffer severity escalations during its life cycle.
- Predicting, at the time of a ticket creation, the maximum severity that the ticket could take reach during its life cycle. This prediction provides valuable information since from the very beginning the real scope of a ticket can be known.
- Predicting, in case an escalation is going to happen, when it will occur.
- Predicting the value of the new severity. This value is very important because a different treatment must be given when a ticket is escalated from 1 to 2 that when it goes from 1 to 5. If this parameter can be correctly predicted, the prioritization of ticket resolution can be highly improved.

5.1. Training and Classification

Each of the sub-problems proposed above is solved independently because each of them requires a different division of tickets in order to obtain the best results. E.g., while the problem of predicting whether there will be an escalation in a ticket or not, does not require any division. Nevertheless, in a more complex problem like predicting the new severity those divisions are essential.

When dealing with the first problem, the only information for predicting whether the ticket will suffer an escalation or not, is the one available at the creation moment. Knowing that a ticket will suffer one or more escalations during its life, although it is not a relevant fact by itself, will allow studying other parameters such as: prediction of maximum severity it will reach, the moment it will happen, etc., as explained later. Given the simplicity of this problem, the results obtained with a C4.5 tree algorithm are close to 100% correct, without the need of applying complicated transformations on data.

Once the first problem is solved, we get two groups of tickets: those that will escalate at some point and those that will not. The next three subproblems only affect those tickets that will supposedly suffer escalations. To solve the problem of predicting the highest severity, the study starts with the available information about a ticket at its creation moment, just in the same way as in the previous problems. The evolution of a ticket does not always follow the same patterns; therefore, these results are just tentative. Besides, they are useful in order to distinguish which tickets must be first processed because of their tendency to produce serious consequences, and which ones are less critical.

Next, the results of applying algorithms from different families of machine learning techniques are shown. Given that these tests have been applied to each different network (this is, as it has already been explained, because of the different casuistic of each of them), the results below refer to just one of the networks, as a representative sample of the results obtained for all of them.

	Tickets created automatically		Tickets created manually	
	Acc.(%)	MAE	Acc.(%)	MAE
Bayes Net	97.1	0.012	59.9	0.125
Naive Bayes	97.1	0.011	54.5	0.142
C4.5	97.2	0.013	61.9	0.131
Decision Table	97.2	0.015	60.2	0.1427

Figure 3. Results for the problem of predicting the maximum escalation: accuracy (Acc.) and mean absolute error (MAE)

Every time an action is performed over a ticket, it is possible that the assigned severity changes because it does not correspond with the real severity any more. That is the reason why it is interesting being able to predict these changes. As in previous cases, the table below shows the results obtained for one of the networks, as a relevant sample of the overall results.

	Correctly classified	Mean absolute error
Bayes Net	87.4	0.050
Naive Bayes	88.0	0.05
C4.5	92.4	0.029
Hyper Pipes	45.1	0.240
Decision Table	91.6	0.037

Figure 4. Results for the problem of predicting the new severity after an escalation

Sometimes, it is useful to know the moment when a severity escalation will occur. Even more interesting than knowing the exact moment with a higher error, it is preferred a more accurate result distinguishing between three intervals: escalations in less than an hour, escalations occurred between one and five hours and escalations after more than five hours. The next table shows the results for the same network as in the previous section.

	Correctly classified	Mean absolute error
Bayes Net	91.7	0.0461
Naïve Bayes	91.0	0.0479
C4.5	92.2	0.0514
Decision Stump	88.4	0.0627
Hyper Pipes	86.7	0.331
Decision Table	92.4	0.0587

Figure 5. Results for the problem of predicting the elapsed time before the escalation

According to the results obtained, we can conclude that the best algorithm for predicting if the escalation will occur and the maximum value of severity, as well as predicting the new escalation value, is C4.5.

For the problem of predicting how long will the ticket take to escalate, even though the results for C4.5 and Decision Table are quite similar, the former has been selected due to the lower mean absolute error.

6. Results

Although the project is still in a development phase, the preliminary results and the improvement in Trouble Ticketing process management already offers promising results: on the one hand, the possibility of predicting on-site works with a reliability over 94% allows a big company planning in advance the routes and fees of technical specialists, which redounds to a decrease in the resolution time, as well as in the number of necessary technical staff to deal with the same number of tickets. On the other hand, the problems that have been solved about the prediction of the severity increase have allowed ticket prioritization to stop being a task performed only by expert people, becoming an automated task, getting better resolution times in critical tickets and a considerable improvement in the decrease of critical cases, since dangerous tickets that used to be unnoticed because of its low initial severity are now being attended at the right moment.

In Figure 6 it is represented an estimation of the number of hours that would have been saved if the system had been already operative during the last month, where it is compared the time needed when a technician has to decide manually whether a given ticket will need on-site work or not with the time needed my the classifier system.

Tickets needing on-site work	18,020
% Correctly classified by the system	93.75
Average time needed for manual on-site work sending (hours)	13.24
Average time needed for automatic on-site work sending (hours)	0.025
Saved time (hours)	223,241.52

Figure 6. Monthly amount of gained time by the on-site work predictions

Obviously, that amount of hours is not a direct profit, since the people in charge of deciding whether a ticket needs or not on-site work do not actually spend 13.24 hours taking a decision, but developing other tasks. The real profit has to do with the perceived quality; a user with a problem that can not be remotely solved will see how a technician comes almost immediately (with the resources constrain of each moment, of course) instead of waiting for hours until somebody takes a decision.

7. Conclusions

Although this work is still in progress, the preliminary results demonstrate that machine learning activities can be successfully applied in order to optimize the needed resources for solving network failures as managed by Trouble Ticket Systems, contributing to a new and automated approach to a process which has been traditionally solved by means of expert knowledge, which is expensive, scant and fallible.

Machine learning skills, unlike the first approximations based on expert systems, have demonstrated to react correctly with confuse and noisy information like that from Trouble Tickets, created by heterogeneous machines and human technical staff, allowing to capture this information and use it to re-feed the TT-System and improve its operation.

8. Future work

The results obtained in this first approach let many other problems related to the process of Trouble Ticketing appear like good candidates for being solved in a second phase of our work. The most interesting of them, and the one we will try to address in the short term, is the association between tickets that refer to the same problem, or those tickets that derivate from other failures.

This problem entails the search for behaviour patterns in tickets to find those that are not caused by an isolated problem, but are a consequence of another higher level ticket. Solving this problem will allow to drastically decrease the number of tickets that are active at each moment, locating those tickets that refer to the root problem (those higher in the hierarchy) and whose resolution will allow the automatic closing of all those that descend from it. From the point of view of the technicians attending the tickets, rooting them will help reducing the number of open problems and helping them discovering relations between them.

Nevertheless, although this is the short term aim of the system, the number of possible problems to be solved in this field is quite extensive, opening an important line of work in this direction that could be generalized and applied later to a great number of business sectors.

9. References

[1] L. Bernstein, C.M. Yuhas, “Expert systems in network management-the second revolution”, *IEEE Journal*, 1988, pp. 784-787

[2] B. Ericsson, “Experiences from Expert System Projects in Network Management”, *International Conference on Information and Data Communication II*, North-Holland Publishing Co., 1988, pp. 161-165

[3] A. Sen, “INAM – An expert system for intelligent network-alarm management”, *Proceedings of the 8th International Conference on Digital Satellite Communications*, Paris, 1989, pp. 509-519

[4] F. Provost, U. Fallad, “A survey of methods for scaling up inductive algorithms”, *Data Mining and Knowledge Discovery*, 1999, pp. 131-169

[5] A. Makanju, N. Zincir-Heywood, “Investigating Cross-Platform Robustness for Machine Learning Based IDSs on 802.11 Networks”, *International Journal of Computer Science and Network Security*, 2007

[6] R. Sasisekharan, V. Seshadri, and S.M. Weiss, “Data Mining and Forecasting in Large-Scale Telecommunication Networks”, *IEEE Expert*, IEEE Educational Activities Department, New Jersey, USA, 1996, pp. 37-43

[7] D. Thunkaram, H.P. Vijaynarasimha, “Artificial Neural Networks and SupportVector Machine approach for Locating Faults in Radial Distribution Systems”, *IEEE Transactions on Power Delivery, vol 20*, 2005, pp. 710- 721

[8] M. Kawahara, H. Kawano, “An Application of Text Mining: Bibliographic Navigator Powered by Extended Association Rules”, *Hicss*, IEEE Computer Society, Los Alamitos, CA, USA, 2000

[9] G.H. John, R. Kohavi, K. Pfleger, “Irrelevant features and the subset selection problem”, *International Conference on Machine Learning*, 1994, pp. 121-129

[10] R. Kohabi, “Wrappers for feature subset selection”, Artificial Intelligence, 1997, pp. 273-324.