Application of Factor Analysis on Improving Diagnostics of Broadband Services

Boris Spahija, Željko Deljac
HT – Croatian Telecom
Savska 32, Zagreb, Croatia
E-mail: Boris.Spahija@t.ht.hr, Zeljko.Deljac@t.ht.hr

Abstract: **With growing number of customers and a demand for higher bandwidth network connections, dictated by introduction of new and improved services, broadband networks are operating near the highest capacity. This reduces the redundancy and fault tolerance of the network, which influences the quality of the service provided to customers. This paper aims to improve the fault management process by introducing improvements based on factor analysis of archived diagnostics data. By recognizing the cause of the fault it will enable the service providers to dispatch the appropriate technicians and resolve the fault in the first attempt.**

1. INTRODUCTION

Constant development of new and improved broadband services, as well as rising number of customers, is creating a growing demand for higher bandwidth network connections. The industry is responding by introducing new technologies that enable faster data transmission over copper telephone lines, like ADSL2 [1], ADSL2+ [2] and VDSL [3]. This is achieved by utilizing a broader frequency band, which reduces the redundancy and fault tolerance, making them more prone to service faults.

Since service faults directly influence customers' satisfaction, fault management is becoming increasingly important domain for Internet Service Providers (ISP). An efficient and instantaneous response to a detected fault is expected and achieving it is a tendency for ISPs. The optimization of fault management process is also directly linked to the cost of maintaining the network which makes accurate diagnostics of the fault and resolution it in the first attempt a priority.

The amount of reported faults per day dictates a number of employees dedicated to fault management process. The employees are usually divided through a few levels of technical expertise: first level is responsible of contact with customers and has a limited tools at their disposal; second level of service experts with advanced software that enables modifications service configuration from the remote location; and third level of field technicians with the ability to solve the faults at the location of the fault occurrence. A typical fault undergoes an analysis and attempted resolution starting at first level, and the process continues sequentially on second and third level until the fault is successfully solved. This makes the process inefficient by often utilizing technicians that don't have the appropriate tools to solve the fault and it increases the possibility of inaccurate diagnostics. Also, it makes it needlessly more expensive and the prolonged service degradation causes dissatisfaction for the affected customers.

The faults caused by elements in core segments of the network are easily detectable since they affect a larger base of customers and a simple analysis of a common element will usually reveal the faulty object. This is why the faults in the access part of the network and customer premises present the biggest challenge.

To manage the process effectively efficiently, we are proposing a solution that will enable early and automated diagnostics, providing the adequate process steering in the initial stages of the process. Additionally, reliable diagnosis of the faults enables early selection of the appropriate level of technicians, consequently reducing the costs of the service maintenance and the cost of service. This paper presents an analytical approach to solving the problem of wrongly diagnosed faults by automating the diagnostics process. The method is based on factor analysis of the data collected in the database of historical cases.

The overview of factor analysis, including related work, is given in Chapter 2. Chapter 3 describes data collection process and types of input data that have been analyzed, while the results of factor analysis applied to data are presented in Chapter 4. Chapter 5 contains results of correlation of obtained factors with fault causes, with conclusion in Chapter 6.

2. FACTOR ANALYSIS OVERVIEW

Factor analysis originates from Spearman [4] and its basic principle is based on presumption that a smaller number of common factors exist that influences a larger number of surface attributes. The correlation of observed variables is explained by their mutual dependence on latent variables, called factors. There are two basic types of factor analysis: Exploratory and Confirmatory factor analysis. Exploratory is used to explore patterns in the data, while the confirmatory is used to test explicitly stated hypotheses and is usually regarded as a theory-confirming model [5].

The Figure 1 demonstrates the principle of factor analysis. Squares represent observable variables while circles represent latent variables or factors. The one-directional arrows are representing linear influences of causation process, while the two-directional represent correlation between variables [6].

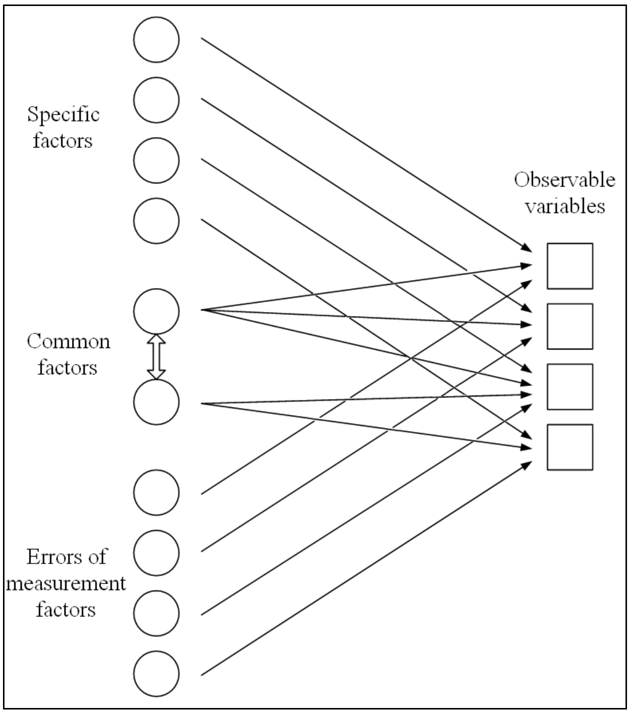

Figure 1 – Factor analysis path diagram

Factor analysis for p observed variables $x_1, \ldots x_p$ with their mean values $\mu_1, \ldots, \mu_p$, written as a linear combination of k unobserved variables $F_1, \ldots, F_k$, and unknown constants l_{ij} (1):

$$X_i = \mu_i + l_{i1}F_i + \ldots + l_{ik}Fk + \varepsilon_i \quad (1)$$

With ε_i representing independently distributed errors, indicating the degree that an instance differs from the average value.

Factor analysis can be written as a matrix equation (2):

$$x = \mu + LF + \varepsilon \quad (2)$$

Where x is a $p \times n$ matrix of observed variables, L a $p \times k$ matrix of factor loadings and F a $k \times n$ matrix of unobserved variables with n being the number of observations, p the number of observed variables and k the number of unobserved variables or factors. Using the known data on the left side of the equation, factor analysis derives factors F and constants l, known as loadings. Loadings measure the strength of the relation between each factor and the observed variables. Apart from loadings, factor analysis also creates scores for each data instance, called factor scores - determining linear mathematical relationship between generated and observed data.

Another way to specify factor analysis is (3):

$$cov(x) = \Lambda\Lambda^T + cov(\varepsilon) \quad (3)$$

The objective is to clarify the relations between observed parameters, as well as linking it with originating factor – the cause of fault. Since we are targeting a specific part of the network topology, the hypothesis is that the number of possible factors affecting the network parameters is equal to the number of functional elements that build the network.

Factor analysis is most often applied to problems in psychometrics, marketing, economics, geochemistry, ecology and hydrochemistry. Voudouris et al. [7] applied factor analysis in studying the factors that affect groundwater quality, determining the importance of hydrogeological parameters and their correlation, similarly Boyacioglu et al [8] used it to assess the quality of surface water. Rahman et al [9] applied it to aerosol composition data to identify fingerprint of the most significant factors. It has also been applied to fault diagnosis in railway track circuits by Come et al. [10]. In medicine, Wu et al [11] investigated the risk variables of metabolic syndrome with factor analysis.

3. BROADBAND DIAGNOSTICS DATA

The data for the analysis is collected from all the diagnostics systems monitoring the status of ADSL line, DSLAM (Digital Subscriber Line Access Multiplexer) and CPE (Customer-Premises Equipment) equipment available to network operators [11]. The parameters and their description are listed in Table 1.

Parameter *Line Attenuation* describes the current state of a line, while variables 1-8 are describing errors that have been collected of a line in the previous 24 hours. The listed variables are being collected for both upstream and downstream, making a total of 15 variables.

Since the cause of the fault is assessed by field technicians, a certain percentage of faults aren't diagnosed accurately. This causes noise in the data and can lead to a misleading results of analysis. This is why it was important to filter the data by removing the faults that might be diagnosed incorrectly. The lines included in the analysis have been monitored for a period of 30 days after the reported fault has been cleared, to ensure that the repair has been made on the correct network element.

Lines that have had a second fault reported in the monitored period have been excluded from the analysis. This still leaves a possibility of errors in data collection process, since the input had to be collected from numerous sources and can suffer from errors in subjective assessment in case of faults with more suspected faulty elements.

After removing lines with a repeated fault in the monitored period, 700 lines have been included in the analysis. Parameters LOFS, LOSS, LOLS, ESS, SESL and Line Attenuation have been collected for both upstream and downstream direction.

Parameter	Description
Number of resynchronisation	Number of resynchronisation in the previous day
Loss of framing seconds (LOFS)	Number of seconds in the previous day with at least one loss of framing error
Loss of signal seconds (LOSS)	Number of seconds in the previous day with at least one loss of signal error
Loss of link seconds (LOLS)	Number of seconds in the previous day with at least one loss of link error
Errored seconds (ESS)	Number of seconds in the previous day that contained one or more bit errors
Severely errored seconds (SESL)	Number of seconds in the previous day that contained more than 30% of errored blocks
UASL	Number of seconds with the interface unavailable
INITS	Number of line initializations
Line Attenuation	Attenuation of the ADSL line at ADSL2+ standard

Table 1- Collected Parameters

Most common causes of faults are: DSLAM port, Telephone copper pair, ADSL splitter, Home installation and ADSL modem. Apart from failure of network path elements, fault can also be caused by a configuration error or a transient effecting Customer premises equipment or DSLAM-s.

4. FACTOR ANALYSIS APPLIED

In this paper we are proposing the application of a latent variable model in order to improve the precision of broadband faults diagnostics. The applied method, factor analysis, explains a set of observed set of continuous variables by a linear relation of a smaller set of continuous variables.

Since there is no firm hypothesis about the relationship of observable variables and latent factors, we will perform exploratory factor analysis with the number of factors based on the number of elements that can be identified as a cause of the fault. As evident from Figure 2 possible fault locations in the access part of the network are modem and ADSL splitter at the customers premises, customer wiring, local loop copper wires located between customers and a central office and DSLAM ports.

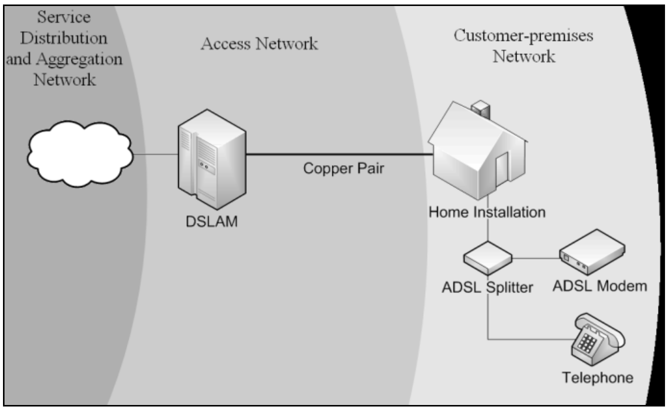

Figure 2 – Network architecture

The number of factors has been determined by analyzing the topology of the network and the possible influences affecting fault occurrence.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Var1	0.3077	0.2493	0.7610	0.5085	-0.0200
Var2	0.3152	0.0681	0.4789	0.2838	0.5703
Var3	0.9834	-0.1024	-0.1285	-0.0558	-0.0028
Var4	0.9824	-0.1030	-0.1351	-0.0576	-0.0064
Var5	0.2322	0.0802	0.2598	0.1981	0.4197
Var6	0.0788	0.4688	0.1024	-0.1585	0.0777
Var7	0.3640	0.0547	0.2756	0.2337	0.7848
Var8	0.1219	0.8220	0.1942	-0.5165	-0.0022
Var9	0.4846	0.1921	0.5548	0.3302	-0.1595
Var10	0.3391	0.1509	0.4297	0.3146	-0.0441
Var11	-0.0345	-0.1114	-0.0523	0.0233	-0.0346
Var12	0.0603	0.2526	-0.1326	0.3887	0.0114
Var13	0.3078	0.2441	0.2679	0.3987	0.0582
Var14	0.1113	0.8051	0.1598	-0.4877	0.0600

Table 2 – Factor loadings

By applying factor analysis to the collected data, factor loadings were extracted, as shown in Table 2. Figure 3 shows factor scores for a sample of 250 faults with each line representing one of 5 selected factors. A dominant factor for each reported fault can be identified by the highest score affecting the measured parameters.

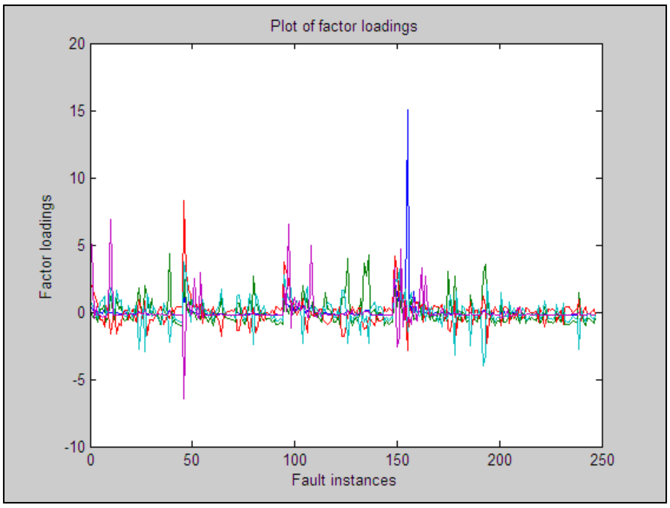

Figure 3- Factor scores

5. CORRELATION OF RESULTS WITH THE REAL DATA AND RESULTS EVALUATION

Our assumption is that each fault cause generates a specific footprint in observed variables and therefore can be associated with a factor calculated using factor analysis.

When we compare fault cause estimations made by technicians with the dominant factor for the same faults, the resulting percentage of correspondence is presented in Table 3.

	Cause1	Cause2	Cause3	Cause4	Cause5
Factor1	0%	0%	0%	79%	21%
Factor2	19%	5%	54%	18%	5%
Factor3	16%	22%	16%	8%	38%
Factor4	44%	10%	15%	15%	16%
Factor5	13%	36%	15%	29%	7%

Table 3 – Correspondence of factors and fault causes

The dissipation of results can be explained by noise interference in the data collection process. Since the data has been collected by various technicians, differences in assessment can occur. This is partially caused by the complexity of broadband service delivery, where the quality of a service is determined by the quality of components in the network path. Often a swap of an element can improve service quality of marginally degradated lines enough to allow service functionality, even though the swapped element isn't the main cause of the degradation. Also, more complex elements can have various subcomponents brake down, which can only be determined by detailed examination of the faulty element.

Figure 4 shows the comparison of diagnostics accuracy of second level technicians and the diagnostics made by the system described in the paper. The comparison is based on a sample of 250 reported faults and is verified by field technicians.

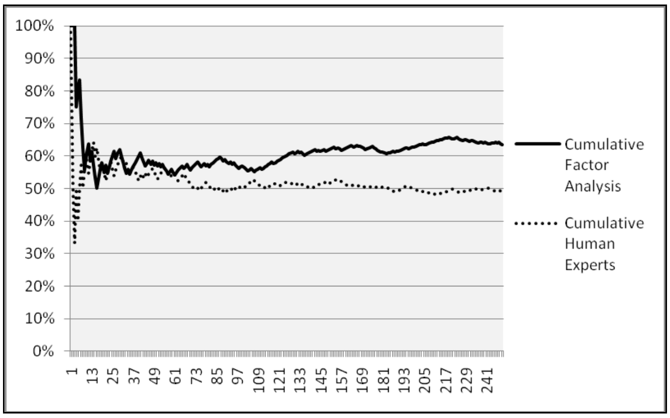

Figure 4- Accuracy comparison of factor analysis and human experts

As evident from the Figure 4, the system based on factor analysis manages to perform with a higher cumulative accuracy than a human expert on a remote location.

6. CONCLUSION

In this paper it has been demonstrated that broadband diagnostics can be performed by a system based on factor analysis on a level similar or better in accuracy then diagnostics made by human technicians. By applying factor analysis to archived diagnostics data, we were able to partially differentiate fault causes, which enabled early diagnostics with a satisfactory accuracy.

The method has shown potential for further development which should be focused on enhancing input data quality by reducing the noise in the data under analysis. Ideally, it should be conducted in a laboratory environment, ensuring the uniformity of extrinsic influences.

REFERENCES

[1] ITU-T G.992.3, "Asymmetric digital subscriber line transceivers 2 (ADSL2)", 2009

[2] ITU-T G992.5, „Asymmetric digital subscriber line (ADSL) transceivers – Extended bandwidth ADSL2 (ADSL2plus)", 2010

[3] ITU-T G993.1, „Very high speed digital subscriber line transceivers (VDSL)", 2004
[4] Spearman, C., "General intelligence objectively determined and measured", American Journal of Psychology, 1904
[5] Hurley, A.E., Scandura, T.A., Schriesheim, C.A., Brannick, M.T., Seers, A., Vandeverg, R.J., Williams, L.J., "Exploratory and confirmatory factor analysis: guidelines, issues and alternatives", 1997
[6] Tucker, L.R., MacCallum, R.C., „Exploratory Factor Analysis", 1997
[7] Voudouris, K.S., Lambrakis, N.J., Papatheothorou, G., Daskalaki, P., "An Application of Factor Analysis for the study of the Hydroglelogical Conditions in Plio-Pleistocene Aquifers of NW Achaia", 1997
[8] Boyacioglu, H., Hayal, B., Orhan, G., "Application of Factor Analysis in the Assessment of Surface Water Quality in Buyuk Menderes River Basin", 2005
[9] Rahman, S.A., Hamzah, M.S., Wood, A.K., Suhaimi, E., Salim, N.A.A., Sanuri, E., "The Application of Factor Analysis to Identify Sources in Klang Valley Aerosol", 2009
[10] Come, E., Oukhellou, L, Aknin, P., Denoeux, T., "Independent Factor Analysis with mixing constraints in a semi-supervised framework, application to railway device fault diagnosis", 2009
[11] RFC 2662, "Definitions of Managed Objects for the ADSL Lines", http://www.ietf.org/rfc/rfc2662.txt , 1999
[12] Timmerman, M.E., "Factor Analysis", 2005