

# Causal Risk Models: A proposal based on associative classes<sup>1</sup>

Modelli di Rischio Causali: Una proposta basata su classi associative

Paola Cerchiello

Department Of Statistics 'L. Lenti', University of Pavia, [paola.cerchiello@unipv.it](mailto:paola.cerchiello@unipv.it)

Paolo Giudici

Department Of Statistics 'L. Lenti', University of Pavia, [giudici@unipv.it](mailto:giudici@unipv.it)

**Riassunto:** nell'ambito della gestione dei rischi connessi all'attività di una società di telecomunicazioni, ruolo rilevante è assunto dalle problematiche rientranti nella categoria di rischio operativo. Con tale espressione ci si riferisce a tutte le perdite o gli eventi dannosi dovuti a cause rientranti nelle quattro macrocategorie individuate da Basilea: persone, processi, sistemi, eventi esterni. La stima di modelli causali in grado di mettere in relazione le possibili cause con le perdite (su scala ordinale) risulta essenziale nell'ambito di un piano di gestione, mitigazione e prevenzione di rischi operativi. In questo contesto proponiamo l'impiego di regole associative, comunemente utilizzate in ambiti differenti (tipicamente MBA), valutandone la validità su un database reale fornito da una società di telecomunicazioni.

**Keywords:** IT-Operational Risk, Association Rules, HyperLift.

## Introduction

According to Basel Committee and common industry, Operational Risk can be defined as 'the risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events'. In particular we focus on IT-intensive organizations, in other words great relevance is given to losses coming from processes and systems intrinsically tied up to the core business of Telco enterprises.

The main objective is to associate the noticed causes of a specific occurred problems (interruptions) to either the resulting loss or the customer type, in order to reveal regularity and associative patterns between those kind of data.

In this context we propose to employ a typical unsupervised model, the association rules, developed in the field of computer science and typically used in other fields like Market Basket Analysis and Web Mining. To apply the association rules, the data can present either a transactional or binary format. The employed R package 'ARules' allows to choose between both of them and to transform one structure into the alternative one.

The available dataset is provided by an israeli company providing Value Added Services (VAS) to SMEs, basically, a set of communications services, both voice and data. Those data deal with the IT problems risible from a communications service

---

<sup>1</sup> This Paper is financially supported by MUSING project (FP6 /027097) and Italian National grant FIRB

provider. In particular we analyze a database compiled by technicians called to solve the clients' IT-problems. Moreover we pay attention to 3 main variables:

- **Cause:** it is a qualitative variable containing the wording of the problem occurred to the client. For example: 'Power Up', 'Reset\_Ld', 'Card' etc.
- **Severity:** it is a quantitative ordinal variable referring the impact of the loss in term of severity: 1 'high', 2 'medium', 3 'low'.
- **Customer type:** it is a qualitative variable referring the typology of client according to a classification similar to Basel Committee one. For example: 'Banking', 'Hotel', 'Defence' etc.

In order to evaluate the presence of a possible hierarchy among the different causes, the analysis has been carried out employing the 'Cause' variable on different generalization levels (5 macro categories, 31 categories, complete original granularity).

The obtained results have been compared in terms not only of the classical interesting measures –Support, Confidence, Lift- but also considering a more robust version of such indexes like HyperLift .

| #RULES                 | lhs                        | rhs                   | Support    | Confidence | Lift      | HyperLift | Phi          |
|------------------------|----------------------------|-----------------------|------------|------------|-----------|-----------|--------------|
| 1                      | {Prob.Code_rid=SEC08} =>   | {Severity=MEDIUM}     | 0,0667675  | 1          | 5,3195402 | 4,1756757 | 0,555912448  |
| 2                      | {Severity=MEDIUM}          | {Prob.Code_rid=SEC08} | 0,0667675  | 0,3551724  | 5,3195402 | 4,1756757 | 0,555912448  |
| 3                      | {Prob.Code_rid=SEC01} =>   | {Severity=LOW}        | 0,09896283 | 1          | 1,399879  | 1,316092  | 0,209569952  |
| 4                      | {Prob.Code_rid=INT04} =>   | {Severity=LOW}        | 0,06633535 | 1          | 1,399879  | 1,2953586 | 0,16855489   |
| 5                      | {Prob.Code_rid=SEC02} =>   | {Severity=LOW}        | 0,05920484 | 1          | 1,399879  | 1,2924528 | 0,15863367   |
| 6                      | {Prob.Code_rid=NTC01} =>   | {Severity=LOW}        | 0,05164218 | 1          | 1,399879  | 1,2849462 | 0,147564015  |
| 7                      | {Business.Line=Defense} => | {Severity=LOW}        | 0,10306828 | 0,7923588  | 1,1092065 | 1,0506608 | 0,06677984   |
| 8                      | {Business.Line=Banking} => | {Severity=LOW}        | 0,12705272 | 0,7188264  | 1,00627   | 0,9623568 | 0,004594276  |
| 9                      | {Business.Line=Health} =>  | {Severity=LOW}        | 0,14282627 | 0,7130529  | 0,9981877 | 0,9565847 | -0,001434298 |
| 10                     | {Business.Line=General} => | {Severity=LOW}        | 0,05747623 | 0,7247956  | 1,0146262 | 0,9466192 | 0,006788042  |
| MINIMUM SUPPORT = 0.05 |                            |                       |            |            |           |           |              |
| # DISCOVERED RULES =10 |                            |                       |            |            |           |           |              |
| POSITIVE               |                            |                       |            |            |           |           |              |
| INDIFFERENCE           |                            |                       |            |            |           |           |              |
| NEGATIVE               |                            |                       |            |            |           |           |              |

**Figure 1:** Extract of some discovered association rules

## References

- Agrawal R., Imieliński T., and Swami A. (1993). Mining Association Rules between Sets of Items in Large Databases. Proc. Conf. on Management of Data, 207–216. ACM Press, New York, NY, USA.
- Giudici P., (2003). Applied data mining. Wiley.
- Hahsler M., Hornik K., and Reutterer T. (2005). Implications of probabilistic data modeling for rule mining. Report 14, Research Report Series, Department of Statistics and Mathematics, Wirtschaftsuniversität Wien, Augasse 2-6, 1090 Wien, Austria.
- Omiecinski E. R. (2003). Alternative interest measures for mining associations in databases. IEEE Transactions on Knowledge and Data Engineering, 15(1):57–69, Jan/Feb.
- Tan P.-N., Kumar V., and Srivastava J. (2004). Selecting the right objective measure for Association analysis. Information Systems, 29(4):293–313.