

Cox regression based process optimization

János Baumgartner, Zoltán Süle, János Abonyi

Data mining is an efficient tool to reveal information and to discover correlations derived out of this data. Process mining and event analysis become more and more popular research area for investigating a sequence of events. Examining the available set of information leads straight to the opportunity of either iterative process development or optimization. According to the general interpretation of survival analysis [1][2] a process can be investigated by focusing on a special event of interest, thus an estimation can be given for the expected duration of surviving time. The Cox's proportional hazard model is capable to divide the entire investigation period into spells (i.e. sub periods) [3]. Thus, the possibility for investigation of different sub periods can be ensured, so the entire process is influenced by each sub process in every well-defined time slot, hence the overall risk and the shape of the survival or hazard function also differs from time to time. We introduce a novel methodology by which time and also cost can be saved by determining the optimal sequence of sub processes in the considered process. Contrary to the classical survival analysis the core idea is to examine the data of a test process consisting of sub process steps, and based on the gained information the sequence of these sub elements can be redesigned. Additionally, when parameters also have to be taken into account, the result of the investigation is affected accordingly. Therefore, parametrical survival patterns can be fitted to the problem, so the risks for each time period can be determined. Using Cox regression we can highlight those process steps including those relevant parameters which increase significantly the risk. It is also important to emphasize that the fault of a process step not exclusively means the fault of the entire process, it assumes a rising risk of the overall fault. These individual sub hazard functions assigned to the different time periods build up a complex characteristic survival function of the regarded process. As a further development we inspect the possibilities of extending the basements as a problem class in connection with neural networks. The results are illustrated through a realistic example taken from manufacturing and analysis of education data.

Acknowledgements

This publication has been supported by the Hungarian Government through the project VKSZ_14-1-2015-0190 - Development of model based decision support system for cost and energy management of electronic assembly processes. The research of Janos Abonyi has been supported by the National Research, Development and Innovation Office - NKFIH, through the project OTKA - 116674 (Process mining and deep learning in the natural sciences and process development).

References

- [1] Brenner, H., Castro, F., Eberle, A., Emrich, K., Holleczeck, B., Katalinic, A., Jansen, L. (2016), *Death certificate only proportions should be age adjusted in studies comparing cancer survival across populations and over time*, European Journal of Cancer, 52, 102-108.
- [2] Ross, A., Matuszyk, A., Thomas, L. (2010), *Application of survival analysis to cash flow modelling for mortgage products*, OR Insight, 23, 1-14
- [3] Willett, J., Singer, J., (1993), *Investigating Onset, Cessation, Relapse, and Recovery: Why You Should, and How You Can, Use Discrete-Time Survival Analysis to Examine Event Occurrence*, Journal of Consulting and Clinical Psychology, 952-965