Fuzzy Logic for Data Analysis in Big Data Management

Michael Kaufmann

Lucerne School of Information Technology, Zug-Rotkreuz, Switzerland
m.kaufmann@hslu.ch

Abstract. The digital age has fostered the data explosion in which the global information capacity doubles every 3 years (Hilbert and López, 2011). With this speed of growth, a *data intelligence gap* is created: The big data available to an organization is growing exponentially, while the percentage of the data that an organization can process and actually use declines as rapidly (Zikopoulos and Eaton, 2011). Ultimately, this percentage could become infinitesimally small. It is said that big data is the oil of the 21st century and, thus, those individuals, companies, and even nations who possess the skills to turn raw data into something valuable will have major competitive advantages. Therefore, there is pressure for enterprises to adapt and implement a big data management (BDM) strategy, even for companies that are not experienced in this field. However, often the question is not how to implement scalable architectures, but how to get started with big data management in the first place – especially in non-technical companies.

This tutorial has two parts. In the first part, the term big data and its implications are analyzed, and the big data management reference model of Kaufmann et al. (2017) is introduced. This model, the big data management canvas (BDMC) enables practitioners to define and implement big data strategies and applications. The proposed framework extends the existing NIST Big Data Interoperability Framework to make it more actionable by providing a frame of reference for extracting value from big data, called "*data effectuation*". This is accomplished by a knowledge-based embedding of big data management in a frame called "*data intelligence"* and by aligning technical aspects of big data with business aspects.

In the second part, a specific fuzzy management method for applied data science is introduced. Inductive Fuzzy Classification is the data mining algorithm of Kaufmann et al. (2015). This method has been implemented in software (IFC-Filter) and tested in real use cases. The qualitative and empirical evaluation showed its potential especially in communicating mining results to human users. The IFC-Filter supports bottom-up data intelligence by transforming patterns in the data into visualizations of fuzzy set membership functions, thus making these patterns readily understandable for human decision makers.

Keywords. Big Data, Data Science, Inductive Fuzzy Classification

References

Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, *332*(6025), 60–65.

Kaufmann, M., Meier, A., & Stoffel, K. (2015). IFC-Filter: Membership function generation for inductive fuzzy classification. *Expert Systems with Applications*, *42*(21), 8369–8379.

Kaufmann, M., Eljasik-Swoboda, T., Nawroth, C., Berwind, K., Bornschlegl, M., Hemmje, M. (2017). Modeling and Qualitative Evaluation of a Management Canvas for Big Data Applications. Accepted for publication as a regular paper at the 6th international Conference on Data Science, Technology and Applications DATA 2017, Madrid, 24 - 26 July 2017.

Zikopoulos, P., & Eaton, C. (2011). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data* (1st ed.). McGraw-Hill Osborne Media.