

The article "The Data Problem in Data Mining" by Albrecht Zimmermann addresses significant issues in the field of data mining, particularly in the context of pattern mining. Zimmermann identifies three main problems:

1. **Quantitative Performance of Algorithms:** There is a lack of understanding regarding how well pattern mining algorithms perform quantitatively. This includes how effective they are and what their limitations might be.
2. **Choosing Parameter Settings:** Selecting appropriate parameter settings for these algorithms is a challenge. There is little guidance on how to set these parameters optimally, which can lead to inefficient or ineffective mining results.
3. **Relating Patterns to Data-Generating Processes:** It is often unclear how the patterns identified by these algorithms relate to the processes that generated the data. This disconnect can limit the practical applicability of the discovered patterns.

Zimmermann argues that these issues are compounded by a lack of diverse and controlled data sets, which hinders the evaluation and comparison of different algorithms. He suggests that progress in the field requires a scientific approach to digital data generation to create varied data sets with known properties for better algorithm assessment.

The issues Zimmermann raises are related to all aspects of data mining: the data itself, the applications of the algorithms, the knowledge generated from the data, and the techniques used to mine the data. The root problem appears to be an interplay between the quality and variety of data available and the methodologies employed to process this data effectively.