International J. of Engg. Research & Indu. Appls. (IJERIA).
ISSN 0974-1518, Vol.4, No. IV (November 2011), pp. 83-96

CORRELATION-BASED DETECTION OF ATTRIBUTE OUTLIER FOR DATA CLEANING

YUVIKA SINGHAL, ANUPAMA SHARMA AND RANJIT SINGH

Abstract

Data overload combined with widespread use of automated large-scale analysis and mining result in a rapid depreciation of the data quality. Data cleaning is an emerging domain that aims at improving data quality through the detection and elimination of data artifacts. These data artifacts comprise of errors, discrepancies, redundancies, ambiguities, and incompleteness that hamper the efficacy of analysis or data mining. In this paper, correlations between data entities are exploited to identify data artifacts that existing data cleaning methods fall short of addressing. A data cleaning method is proposed for detecting outliers and duplicates.

Keywords : Outlier detection, Attribute outlier, Data subspaces, Threshold optimization.