

Visual Analytics System for Data Quality Improvement

Authors: Hyein Hong, Sangbong Yoo, yejin jin, Yun Jang

Summary:

The machine learning performance is affected by the model and data quality. However, most users focus on modifying the model rather than improving data quality. Different techniques are used to improve data quality depending on the data condition and the cause of poor data quality. Therefore, improving data quality is a timeconsuming and challenging task for users who do not have sufficient knowledge of data. Visual analytics techniques have been proposed to focus on decision support to improve data quality. However, existing studies do not evaluate the impact of data quality improvement on the performance of machine learning models. In addition, existing studies have limitations in which users have to consider all combinations of data quality improvement processes. In this paper, we propose a novel visual analytics system to manage data quality to improve the performance of machine learning models. The proposed system suggests an optimal quality improvement method and process for data to be appropriately used in machine learning models.