

Investigation of Comparability Over Time of Data and Its Impact on Credit Rating Model

Jiali Du

Graduate School of Symbiotic Systems Sciences and Technology,
Fukushima University
Fukushima city, Japan
s2070033@ipc.fukushima-u.ac.jp

Yanwen Dong

Cluster of Science and Technology
Fukushima University
Fukushima city, Japan
dong@sss.fukushima-u.ac.jp

Abstract

Almost all of statistical models and machine learning models are to be built based on a known dataset (learning dataset) that was collected from past or other regions/domains. The precondition for applying these models to new instances is that the data of the new instances are comparable to the learning dataset. The comparability between the new instances and the learning dataset has also a strong impact on the performance of statistical models and machine learning models. There are mainly three types of comparability: comparability over region (geographical comparability), comparability over other domains and comparability over time. Comparability over time refers to the extent to which statistics are comparable or reconcilable over time. In particular, some social or economic indexes are affected clearly by changes over time of economic or social phenomena, comparability over time is an important aspect to improve the quality of various models in the social and economic fields. As major studies have put the emphasis on analysis methods or algorithms, there are lacks of researches dealing with comparability over time.

This study intends to address the issue of comparability over time, our emphasis is put on (1) proposing a method to assess the comparability over time among datasets collected in different periods; (2) clarifying the impact of comparability over time on statistical models through examining how the performance of statistical models changes when the extent of comparability between the learning dataset and fitting datasets is different. As an example of statistical models, we consider the credit rating problem of Japanese regional banks. We collect financial indicators of Japanese regional banks in 2012 (R2012), 2015 (R2015) and 2017 (R2017) respectively, conduct the following examinations and make several new contributions to data analytics research:

- (1) There are no researches published to assess comparability over time, this study is to demonstrate that the number of variables with significantly different means or/and variances can provide a measure to assess to which extent that the two datasets are comparable.
- (2) We apply linear discriminant analysis (LDA) method to construct two credit rating models based on the R2012 (Model A) and R2015 (Model B) respectively, and then use these two models to obtain the credit rating for each bank in R2017. Through comparing the forecasting results of model A and model B, and examining the extent of the comparability among these three datasets, we can clarify the impact of comparability over time on the performance of credit rating models.
- (3) As a data preprocessing tool, data normalization has been applied widely in various fields. The main goal of data normalization is to guarantee the quality of the data, scale the data in the same range of values for each feature in order to minimize bias for one feature to another. In this study, we propose a

new viewpoint to apply data normalization as a tool of improving comparability over time between the datasets.

Keywords

Comparability over time, Credit rating, Linear discriminant analysis (LDA), Data normalization

Biographies

Jiali Du is now a master student at Graduate School of Symbiotic System Science, Fukushima University. She received a bachelor degree in environmental art design from Beijing Forestry University in 2019. Her research interests include data analytics, data mining and big data applications.

Yanwen Dong is a professor in the cluster of science and technology at Fukushima University. He received his bachelor degree in 1982 and a master degree in 1984 from University of Science and Technology Beijing, China. He also received PhD in 1996 from the Osaka Prefecture University, Japan. He worked in University of Science and Technology Beijing as a lecture from 1984 to 1995 and in Faculty of Economics, Fukushima University as an associate professor from 1997 to 2004. His current research interests include cell production system, production scheduling, data mining and management information system.