Potential bias and discrimination in the health care machine learning algorithm

Abstract:

Machine learning has been widely used in analyzing and interpreting electronic healthcare records. The integration of machine learning could improve healthcare decisional making and support the diagnostic operation. Yet predictive models may also induce unexpected consequences. One of them is indirect discrimination. The machine learning algorithm is known as objectively discover pattern and make prediction. However, data-driven predictive models may end up with discriminate certain groups of people and worsen social disparity. In this study, we review a predictive model for heart transplantation survival probabilities and introduce statistic disciplines to measure discrimination. We also computationally analyze selected measures and discuss measuring procedures. This survey is primarily intended for researchers to utilize data mining and machine learning to develop a non-discrimination predictive model. In addition, practitioners and policymakers would use the study for diagnosing potential discrimination by predictive models.

Intro:

Interpreting and analyzing historical data has always been a crucial part of the healthcare decisional making process. (Shenoy, 21 August 2017) As data suggested, the use of electronic healthcare data has increased dramatically in the last five years. (Benjamin A Goldstein, January 2017) With the increasing pervasiveness of electronic health data, machine learning has undertaken a significant role in data handling and analyzing. (Shenoy, 21 August 2017) The data-driven predictive model can assist physicians and staff members who provide healthcare targeted information with diagnostic support or advanced analytics information that improves clinical decision making or offer diverse treatment options. (Milena A. Gianfrancesco, November 2018)

However, even if the algorithm is fair and well-intentioned, the integration of machine learning may end up with discriminate certain groups of people. (Žliobaitė, 2017) Such bias or discrimination could be unintentional or unexpected yet severe. (Žliobaitė, 2017) The penetration of machine learning in other fields outside healthcare has already aroused people’s concern for social or economic disparities such as legal and justice systems, advertisements, or computer vision. (Milena A. Gianfrancesco, November 2018)