

A Methodology for Direct and Indirect Discrimination Prevention in Data Mining

Authors

Hajian, Sara; Domingo-Ferrer, Josep
Universitat Rovira i Virgili, Tarragona

Abstract

Data mining is an increasingly important technology for extracting useful knowledge hidden in large collections of data. There are, however, negative social perceptions about data mining, among which potential privacy invasion and potential discrimination. The latter consists of unfairly treating people on the basis of their belonging to a specific group. Automated data collection and data mining techniques such as classification rule mining have paved the way to making automated decisions, like loan granting/denial, insurance premium computation, etc. If the training data sets are biased in what regards discriminatory (sensitive) attributes like gender, race, religion, etc., discriminatory decisions may ensue. For this reason, antidiscrimination techniques including discrimination discovery and prevention have been introduced in data mining. Discrimination can be either direct or indirect. Direct discrimination occurs when decisions are made based on sensitive attributes. Indirect discrimination occurs when decisions are made based on nonsensitive attributes which are strongly correlated with biased sensitive ones. In this paper, we tackle discrimination prevention in data mining and propose new techniques applicable for direct or indirect discrimination prevention individually or both at the same time. We discuss how to clean training data sets and outsourced data sets in such a way that direct and/or indirect discriminatory decision rules are converted to legitimate (nondiscriminatory) classification rules. We also propose new metrics to evaluate the utility of the proposed approaches and we compare these approaches. The experimental evaluations demonstrate that the proposed techniques are effective at removing direct and/or indirect discrimination biases in the original data set while preserving data quality.

Published in:

[Knowledge and Data Engineering, IEEE Transactions on](#) (Volume:25 , Issue: 7)

Date of Publication: July 2013

Page(s):

1445 - 1459

ISSN :

1041-4347

Digital Object Identifier :

[10.1109/TKDE.2012.72](https://doi.org/10.1109/TKDE.2012.72)

Product Type:

Journals & Magazines

Date of Publication :

03 abril 2012

Date of Current Version:

21 mayo 2013

Issue Date :

July 2013

Sponsored by :

IEEE Computer Society