[0060] FIG. 10 is graphs of fairness metrics for additional example experiments;

[0061] FIG. 11 is graphs comparing cross-validation grid search analysis of additional example experiments for SLSD and R-SLSD models according to non-limiting embodiments or aspects and an ADM;

[0062] FIGS. 12A and 12B are graphs comparing cross-validation grid search analysis of additional example experiments for SLSD and R-SLSD models according to non-limiting embodiments or aspects and Distributionally robust optimization (DRO) and adversarially re-weighted learning (ARL);

[0063] FIG. 13 is a table showing a summary of datasets for example experiments;

[0064] FIG. 14 includes pseudocode for protected group estimations in SLSD;

[0065] FIG. 15 includes pseudocode for protected group estimation in R-SLSD; and

[0066] FIG. 16 includes pseudocode for debiasing of a downstream model in SLSD and R-SLSD.

DETAILED DESCRIPTION

[0067] For purposes of the description hereinafter, the terms “end,” “upper,” “lower,” “right,” “left,” “vertical,” “horizontal,” “top,” “bottom,” “lateral,” “longitudinal,” and derivatives thereof shall relate to the embodiments as they are oriented in the drawing figures. However, it is to be understood that the present disclosure may assume various alternative variations and step sequences, except where expressly specified to the contrary. It is also to be understood that the specific devices and processes illustrated in the attached drawings, and described in the following specification, are simply exemplary and non-limiting embodiments or aspects of the disclosed subject matter. Hence, specific dimensions and other physical characteristics related to the embodiments or aspects disclosed herein are not to be considered as limiting.

[0068] Some non-limiting embodiments or aspects are described herein in connection with thresholds. As used herein, satisfying a threshold may refer to a value being greater than the threshold, more than the threshold, higher than the threshold, greater than or equal to the threshold, less than the threshold, fewer than the threshold, lower than the threshold, less than or equal to the threshold, equal to the threshold, etc.

[0069] No aspect, component, element, structure, act, step, function, instruction, and/or the like used herein should be construed as critical or essential unless explicitly described as such. Also, as used herein, the articles “a” and “an” are intended to include one or more items and may be used interchangeably with “one or more” and

discussed: each of DRO and ARL often underperform or can even intensify the existing biases of the standard Baseline model. In contrast, SLSD and R-SLSD models according to non-limiting embodiments or aspects consistently mitigate these biases, even when SLSD and R-SLSD are trained for the extreme scenario study with the COMPAS dataset.

[0133] Grid search analysis with supervised ADM. FIG. 11 is graphs comparing cross-validation grid search analysis of additional example experiments for SLSD and R-SLSD models according to non-limiting embodiments or aspects and an ADM. FIG. 11 presents the cross-validation grid search analysis for SLSD, R-SLSD, and ADM on each of the ACSIncome and COMPAS datasets. For ACSIncome, using just 1% of the protected data, R-SLSD's performance closely aligns with that of ADM, which utilizes 100% of the protected data, in both utility and fairness. Particularly, most R-SLSD models demonstrate better fairness in terms of both DPD and DPR compared to the majority of ADM models with a roughly equivalent AUC. Surprisingly, for the COMPAS dataset, a role reversal is observed in performance: R-SLSD outperforms both ADM and SLSD in fairness metrics with a greater sacrifice in AUC, while SLSD's performance is approximately similar to ADM's performance.

[0134] Grid search analysis with DRO and ARL. FIGS. 12A and 12B are graphs comparing cross-validation grid search analysis of additional example experiments for SLSD and R-SLSD models according to non-limiting embodiments or aspects and DRO and ARL. For all datasets, except COMPAS, SLSD outperforms other models in both DPD and DPR metrics, however, with a greater loss in AUC. Conversely, R-SLSD consistently shows a balance between fairness and utility. Specifically, R-SLSD performs similarly to the best predictive model, ARL, in terms of AUC, yet R-SLSD significantly surpasses ARL in fairness metrics - DPD and DPR. Interestingly, the results for the COMPAS dataset show an opposite trend between SLSD and R-SLSD compared to other datasets. However, these results may be attributed on the COMPAS dataset to the extreme scenario study, and viewed as outliers.

[0135] SLSD and R-SLSD according to non-limiting embodiments or aspects are compared with two naive baselines and two state-of-the-art approaches. All methods have the same DNN architecture, optimizer, and activation functions. Below are the implementation details:

[0136] Baseline: This baseline classifier is a straightforward vanilla model with a standard binary cross-entropy loss. It focuses solely on accurate predictions,