International Journal of Engineering Research & Management Technology
ISSN: 2348-4039
November-December-2021 Volume-8, Issue-6
Email: editor@ijermt.org www.ijermt.org

INTERSECTIONALITY AND ALGORITHMIC FAIRNESS: UNDERSTANDING AND ADDRESSING COMPLEX BIASES

Godavari Basavaraj
Research scholar Department of Computer Science
OPJS University Churu (Raj)

Dr.Vijay Pal Singh
Associate Professor Department Of Computer Science
OPJS university Churu. (Raj)

ABSTRACT

When users are provided with reasons for the decisions that are made by algorithms, they have the opportunity to learn more about the elements that influence their findings and identify any biases that may exist. Through the use of various approaches such as rule-based models, interpretable machine learning, and post-hoc explain ability processes, it is possible to simplify the understanding of complicated machine learning models. These techniques may also aid in the identification and elimination of biases. The existence of bias in machine learning systems may be attributed to a wide variety of factors. One of the most significant causes is the presence of biases in the training data that was used in the construction of these systems. Wide variety of problems pertaining to fair distribution, and it provided solutions that are doable in addition to promises that can be verified. In light of the results presented in this thesis, new lines of inquiry have been initiated, and it is anticipated that these discoveries will result in the emergence of additional exciting new concerns about the fairness of algorithmic systems. *In the event that historical data reflects social assumptions or structural inequalities, it is possible for machine learning algorithms to inadvertently pick up on and spread these biases. For instance, if a recruiting algorithm is trained on historical data that exposes gender bias in hiring choices, the model may inadvertently favor one gender over the other, so perpetuating discriminatory practices that have been in place in the past. Increasing the diversity of the training data is the goal of data augmentation, which may be accomplished by either generating fake samples or modifying samples that already exist. A reduction in data bias may be achieved using this method by ensuring that underrepresented groups or characteristics are accurately represented. By using techniques like as under sampling, oversampling, and data synthesis, it is possible to supplement the training data and lessen the amount of bias that is present.*

Keywords: *Intersectionality , Algorithmic , Fairness , Biases , Addressing*

INTRODUCTION

Reducing Prejudice in Machine Learning Algorithms

When it comes to reducing prejudice in artificial intelligence, a variety of techniques have been proposed by both academics and industry personnel. A few examples of these procedures include preprocessing the data, selecting a model, and making judgments once the data has been processed. There are, however, limitations and problems associated with each and every strategy. These include the limited availability of training data

that is both representative and diverse, the challenge of identifying and measuring the many different types of bias, and the danger of compromising accuracy in favor of fairness.

In addition, there are ethical problems about which groups should be prioritized while attempting to reduce prejudice, as well as how to prioritize different types of bias. In spite of these challenges, eliminating bias in artificial intelligence is essential to the development of fair and equitable systems that are beneficial to all individuals and to society as a whole. It is essential to conduct ongoing research and contribute to the development of mitigation methods in order to overcome these challenges and ensure that artificial intelligence systems are used in a fair manner.

Methods of pre-processing to reduce data bias:

1 Data enhancement:

Increasing the diversity of the training data is the goal of data augmentation, which may be accomplished by either generating fake samples or modifying samples that already exist. A reduction in data bias may be achieved using this method by ensuring that underrepresented groups or characteristics are accurately represented. By using techniques like as undersampling, oversampling, and data synthesis, it is possible to supplement the training data and lessen the amount of bias that is present.

2 Methods of sampling:

Sampling approaches assist to lessen the biases that are generated by imbalanced datasets. These strategies include selecting a subset of the training data that is typical of the whole. There is a possibility that the impact of data bias on model performance may be mitigated by the use of techniques like as random oversampling, stratified sampling, and SMOTE (Synthetic Minority Over-sampling Technique). These techniques aim to achieve a more even distribution of classes or characteristics within the training data.

3 Algorithms for preprocessing data:

The input data is modified using algorithms that are used for pre-processing data in order to lessen the impact of biases and improve the fairness of the model. Strategies like as feature scaling, feature selection, and dimensionality reduction may be able to aid in alleviating biases that are caused by duplicated or superfluous features. This is done in order to guarantee that the model focuses on relevant data and reduces the impact of extraneous characteristics.

Techniques for mitigating algorithmic bias at the processing stage:

1 Fairness-conscious machine learning techniques:

During the course of the model training phase, fairness objectives or limits are explicitly included into machine learning algorithms that incorporate fairness. These algorithms ensure that the model's predictions are consistent over a wide range of characteristics or demographic categories by simultaneously optimizing the accuracy of the forecasts and the fairness of the model. Several examples include adversarial debiasing, prejudice removers, and fairness-constrained optimization algorithms. A few instances are listed below.

2 Reduction of bias in model training:

During the process of model training, bias mitigation techniques bring about modifications to the learning procedure in order to reduce the amount of algorithmic bias. All of these measures, including reweighting training samples, utilizing adversarial training, adding fairness limits or penalties, and so on, have the potential to help reduce biases in the decision boundaries of the model and reduce the discrepancies in outcomes that are predicted between groups.

3 Regularization methods:

Regularization techniques like as L1 and L2 regularization, which penalize convoluted models or feature coefficients that contribute disproportionately to biased predictions, may be able to assist in the reduction of algorithmic bias.

By setting restrictions on the complexity of the model, regularization procedures lessen the impact of biased features or patterns. This is accomplished by compelling the model to generate representations of the data that are more robust and generalizable.

OBJECTIVES OF THE STUDY

1. To study on reducing Prejudice in Machine Learning Algorithms
2. To study on techniques for mitigating algorithmic bias at the processing stage

RESEARCH METHOD

In this research method descriptions of authentic approaches for MAB that were driven by the pay-per-click auction for internet advertising. 184, 183, and 2 are the several additional research that explore this particular topic. the crowd sourcing problem is provided as a challenge for the creation of a mechanism that is based on MAB algorithm. The upper confidence bounds, often known as UCBs, are the fundamental building blocks of all approaches that are now in use., the Thompson sampling strategy has shown performance guarantees that are slightly higher than those of previous approaches. An strategy that makes use of Thompson Sampling that was suggested by the authors in serves as the inspiration for our method.It is to the best of our knowledge that these strategies have not been used in the design of any MAB mechanism. Utilizing neural networks, we make an effort to develop a technique that is based on the Thompson sampling approach.

Attributes for one wishes

We say that a mechanism M is allocatively efficient if it selects an agent in each round t throughout the course of the mechanism's operation. When this occurs,

$$I_t(b_t) \in argmax\ W\mu_i - b_{i,t}$$

Out of these p objects, each agent is only interested in obtaining one of them. Regarding the acquisition of any of these p resources, Agent I has a value of vi = θi, provided that these things are homogeneous. There is also the possibility that the goods are distinct or heterogeneous; in this scenario, every agent would provide a different value in order to obtain a variety of things. (vi = (θi1 , θi2 , . . . , θip)). Additionally, in order for these commodities to be dispersed to people who place the most value on them, they need to be allocatively efficient (AE). The real guiding concepts v = (v1, v2 . . . ,vn) The information that the agents have about the objects is derived from the personal data that they have. θ = (θ1, θ2, . . . ,θn) = (θi , θ−i) in addition, the strategic agents

have the ability to flag them as inappropriate. (θ ′ 1 , . . . , θ′ n). There is a possibility that the agents would boast about their evaluations if they are not compensated appropriately. Because of this, we need payment from agent I. on the basis of the values that were supplied.

It is necessary for us to establish a system. An allocation rule, as well as a payment rule, denoted by M = (A,P). A. P. decides on a distribution kind. where the payments are established by P, and K is the collection of all potential allocations between them. With the help of this terminology, we will now explain the qualities that we want a mechanism to include..

DATA ANALYSIS

Groves' Redistribution Mechanism: We would want to do so in order to maintain DSIC and AE while also sharing the surplus among the participants to the greatest extent that is practical. As a result, SBB and DSIC are incompatible with one another. This kind of system is referred to as the Groves redistribution mechanism, or simply the redistribution mechanism. In the process of developing a redistribution system, one of the required steps is the creation of an appropriate rebate function. Our ideal solution would be a rebate function that ensures the highest possible refund or the least possible budget imbalance. Additionally, we would want the redistribution process to include the following qualities in addition to DSIC requirements:

First, the possibility (F). It is essential that the entire amount paid to the agents is equal to or lower than the total amount collected that was collected.

Second, the Individual Who Is Rational (IR). By participating in the mechanism, every agent ought to end up benefiting in a manner that is not detrimental.

The lack of identity. Every agent has the same rebate function, which is denoted by the equation ri() = rj () = r(). This might still result in varied redistribution payments as there is a potential that the function input could be considerably different from one instance to the next.

In the process of building a redistribution mechanism for either homogeneous or heterogeneous goods, we may have a linear or nonlinear rebate function of the following kind, as shown in Here is an example of how this may happen. Within the Groves redistribution mechanism, every deterministic and anonymous rebate function f is considered to be DSIC if and only if,

$$r_i = f(v_1, v_2, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n) \ \forall i \in N$$

where, $v_1 \geq v_2 \geq \ldots \geq v_n$.

It is determined that the linear rebate function is 1. The rebates that are given to an agent will be determined by a linear rebate function if the rebate is a linear combination of all of the bid vectors that are still being used by others. In light of this,

$$r_i(\theta, i) = c_0 + c_1 v_{-i,1} + \ldots + c_{n-1} v_{-i,n-1}.$$

,

In spite of the fact that there may be a family of ways that satisfy the aforementioned characteristics, the objective is to identify the redistribution mechanism that redistributes the greatest proportion of the total VCG

payout. An evaluation of the efficiency of the redistribution mechanism is carried out with the help of the redistribution index, which is established. The redistribution index of a redistribution mechanism is the worst-case percentage of VCG surplus that is distributed among the agents. This index gives an indication of how the surplus is distributed. When put another way,

$$e^{ow} = inf_{\theta:t(\theta)\neq 0} \frac{\sum r_i(\theta_{-i})}{t(\theta)} \quad \text{.....(E. q. 1)}$$

First, we will discuss a redistribution mechanism in Optimal Redistribution Mechanisms. When doing so, we will bear in mind the nomenclature that was discussed before as well as the Green-Laffont Impossibility Theorem throughout our discussion.

In order to maximize, one must maximize the total return that is expected. The authors raised the problem by using in their argument. In the case of heterogeneous objects or with a nonlinear rebate function, the OE aim has not yet been addressed.

Table 1: Optimization problem formulation

	OE	OW
Variables :	$c_0, c_1,....,c_{n-1}$	$e^{ow}, c_0, c_1,....,c_{n-1}$
Maximize :	$\mathbb{E}\sum_{i=1}^{n} r_i$	e^{ow}
Feasibility :	$\sum_{i=1}^{n} r_i$ ¡= t	$\sum_{i=1}^{n} r_i \leq t$
Other constraints		
For worst-case :		$\sum_{i=1}^{n} r_i \geq e^{ow} t$
IR :		$r_i \geq 0$

Optimal worst-case scenario (OW)

If there is a guarantee that the redistribution method will provide the agents with an average higher refund, then it is preferable. In order to evaluate a mechanism, we would take into consideration the worst redistribution index that it guarantees in the absence of distributional information. The following model was presented by the authors, and linear rebate functions were used in order to solve it analytically for a homogeneous setting. Further, they declare that the worst-case optimal mechanism is the best one among any deterministic, anonymous redistribution mechanisms that satisfy DSIC, AE, and F.

This is the case regardless of the method's level of anonymity. An explanation of the optimization issue may be found in the following syntax. The objective of this study is to determine and establish the optimality of a nonlinear redistribution mechanism known as HETERO for heterogeneous situations. It is shown in the following theorem, which is referred to as, that there is no optimal mechanism that has a linear rebate function for a variety of conditions. If a redistribution mechanism is both feasible and individually acceptable, then it is impossible for there to be a linear rebate function that is simultaneously DSIC, deterministic, anonymous, and has a redistribution index that is not zero.

Payments and inputs are arranged in this manner.

When it comes to the arrangement and computation of the VCG payment, the neural network has no influence whatsoever, regardless of whether the situation is homogenous or heterogeneous.

These are the equivalent items. The maximum number of similar items that each agent desires is one of the p items that are offered. These are the rates that are being provided. $\theta \in \mathbb{R}^n$. . The bids are assembled in such a manner that v1 ≥ v2 ≥ . . . ≥ vn. Payment made by Agent I

$$t_i = \begin{cases} v_{p+1} & i \le p \\ 0 & i > p \end{cases}$$

……(e.q. 2)

Hence, $t = pv_{p+1}$.

Heterogeneous Objects. All the p objects are different, each agent will submit his valuation for each of the objects. The bids submitted are θ where θ Items that do not conform to a standard. Every single agent will assign their own value for every single object since every single one is different. The prices that are being provided are θ, where θ is the price. We build a particular ordering among these vectors by taking into account the overall utility of each agent as well as their marginal values for each item by using this information. To tackle the problem of commodities allocation, which is analogous to a weighted graph matching problem, the Hungarian Algorithm is considered to be the most effective solution. Once we have obtained the allocation, which we will refer to as k∗, we proceed to use the VCG payment formula in order to ascertain

the payments $t = \sum_{i \in N} t_i$, We build a particular ordering among these vectors by taking into account the overall utility of each agent as well as their marginal values for each item by using this information.

OE-HO stands for ideal in anticipation of uniform-wearing items. The values are obtained from a random distribution that is uniform, and the inputs are arranged in a matrix structure that is (S × n), where S is the batch size. $U[0, 1]$. The batch size is chosen to be as big as possible for n less than 10, S equal to 10,000, and n equal to 10,000, S equal to 50,000. Following this, we apply the ordering and calculate payments for the input values that have been supplied by utilizing the definitions that are found in. This is followed by the feeding of the linear network model, which is seen in, and the Xavier initialization approach is used to initialize the parameters of the model. Following the application of the objective function (Equation 3.2) to the output of the network, the Adam optimizer is used to update the parameters, maintaining a learning rate of 0.0001 throughout the process.

Table 2: e^{oe} for homogeneous and heterogeneous setting.

n, p	Homogeneous			Heterogeneous	
	OE-HO Theoretical	OE-HO Linear	OE-HO Nonlinear NN	OE-HE Linear NN	OE-HE Nonlinear NN

		NN			
3,1	0.667	0.668	0.835	0.667	0.835
4,1	0.833	0.836	0.916	0.834	0.920
5,1	0.899	0.901	0.961	0.900	0.969
6,1	0.933	0.933	0.973	0.934	0.970
3,2	0.667	0.665	0.839	0.458	0.774
4,2	0.625	0.626	0.862	0.637	0.855
5,2	0.800	0.802	0.897	0.727	0.930
6,2	0.875	0.875	0.935	0.756	0.954
10,1	0.995	0.996	0.995	0.995	0.995
10,3	0.943	0.945	0.976	0.779	0.923
10,5	0.880	0.880	0.947	0.791	0.897
10,7	0.943	0.944	0.976	0.781	0.857
10,9	0.995	0.997	0.996	0.681	0.720

Training the nonlinear model is accomplished in a manner that is analogous. We used a total of one thousand nodes in the hidden layer, and the network was trained at a learning rate of ten to the power of four.

When the worst-case scenario occurs, optimal for homogeneous objects (OW-HO) This is done using the linear network model, and the procedure is exactly the same as it was in the OE example. Following the sorting of the input in accordance with the instructions provided in and the computation of the payments, the input is then delivered into the linear network. When the redistribution index achieves its optimal value, the objective is optimized with the learning rate set to 0.0001, and the training process is maintained until the loss decreases and hits saturation. This procedure is repeated until the goal is optimized. In this particular situation, we did not make use of a nonlinear model since, as we discussed in for homogeneous settings, linear rebate functions, furthermore known as DSIC and AE, are the most effective deterministic functions that are now accessible.

In anticipation of non-uniform items, ideal in anticipation (OE-HE) For the second time, a uniform distribution U is used in order to randomly sample the inputs again. The matrix that is being entered has the format of (S × n × p). In the next step, the inputs are grouped in the way that is described in. In the expectation mechanism, the two networks that are applied to identify the optimal situation are shown in respectively. The parameters of the network are initialized using the Xavier initialization method, much as in the case of the homogeneous network. For the purpose of doing the payment calculation, we make use of the scipy library for linear sum assignment. Additionally, we negate the bids before sending them to the function since we want the valuation to be maximized in line with AE. This library allocates objects in a manner that minimizes costs; however, we do this before handing them to the function. Furthermore, in order to make the input matrix square, we add

fake agents or dummy objects with zero value to it. This is because the Hungarian technique for assignment is exclusive to situations in which the number of objects to be assigned is equal to the number of agents. The optimization of the objective function is performed by using the Adam optimizer with a learning rate of 10e − 4 for both the linear and nonlinear models. A constant value of one thousand was assigned to the hidden layer node count in the nonlinear network.

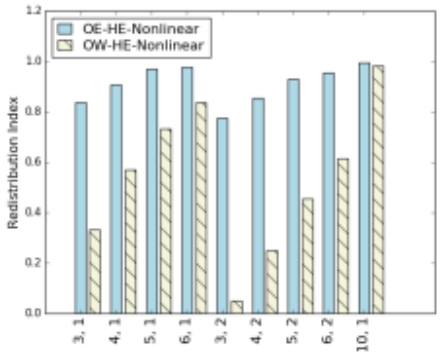

Figure 1: OE-HE-Nonlinear Vs OW-HE-Nonlinear

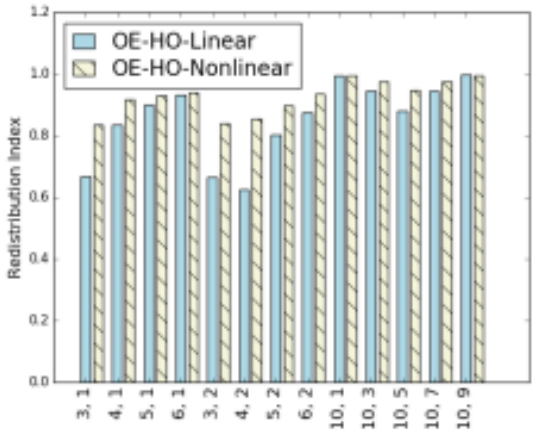

Figure 2: OE-HO-Linear vs OE-HO-Nonlinear

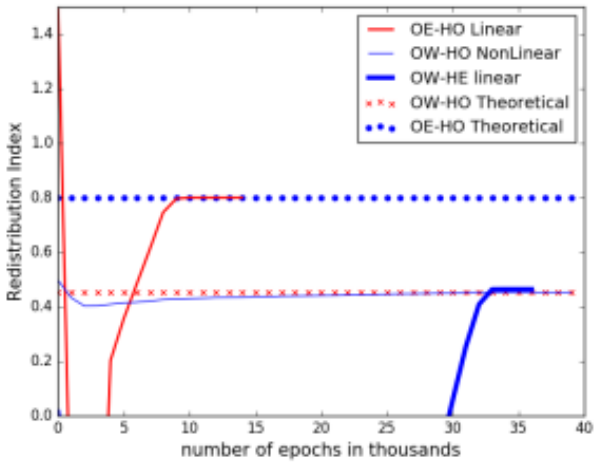

Figure 3 The values of the RI for n = 5 and p = 2 with an epoch change are shown in. Particular training factors are also shown.

- The Tensor Flow library was used in each and every implementation that we carried out. We made use of a GeForce GTX Titan X and a Tesla K40c graphics processing unit (GPU). The amount of time required for network training might vary anywhere from a few minutes to a whole day, depending on the values of n and p.

- It is recommended that the ideal number of input samples (T) for binary settings be two np, but for real value situations, a much higher number is required. Therefore, when the number of samples is less than 13, we use 10,000 samples, when the number of samples is between 13 and 16, we use 70,000 samples, and when the number of samples is between 16 and 100,000, we use 100,000 samples.

- In all experiments, a maximum of 400000 epochs are used, and the constant ρ is set at 1000. This value is stated in the overall loss functions, which can be found in Equations 1 and 2. The network shown in Figure 2 includes a hidden layer that has a thousand nodes, according to an ideal scenario.

CONCLUSION

In a nutshell, addressed a wide variety of problems pertaining to fair distribution, and it provided solutions that are doable in addition to promises that can be verified. In light of the results presented in this thesis, new lines of inquiry have been initiated, and it is anticipated that these discoveries will result in the emergence of additional exciting new concerns about the fairness of algorithmic systems. In a nutshell, we show that neural networks are capable of learning effective redistribution mechanisms, provided that they are given the suitable initialization and a sufficiently defined ordering across valuation profiles. Our investigation has shown that it is feasible to develop nonlinear rebate functions for homogeneous scenarios that are superior than optimal expectation linear rebate functions in terms of performance results. When dealing with heterogeneous items, it is possible that we will construct optimal in expectation rebate functions, which cannot be addressed theoretically. There are a great deal of challenges that need to be conquered here.

REFERENCES

[1] Dastin, J., & Bartz, D. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters.com.

[2] National Transportation Safety Board (2019). Preliminary Report Highway HWY19MH010 [PDF file].

[3] European Commission. (2018). General Data Protection Regulation (GDPR).

[4] New York City Commission on Human Rights (2021). Legal Enforcement Guidance on Discrimination Based on Disability: Reasonable Accommodation under NYC Human Rights Law [PDF file].

[5] Crawford K., Calo R., Barocas S., Bechmann A., & Gillespie T. (2019).

[6] Crawford K., Calo R., Barocas S., Bechmann A., & Gillespie T. (2019).

[7] Smith-Lovin L., Douglas C., & McPherson J.M (2001). "You Are Who You Know: A Network Approach To Gender." American Sociological Review, 66(6), 918-937.

[8] Smith-Lewis A., & Wyner A.Z (2020). Reducing Bias in Artificial Intelligence Systems: A Systematic Review. Journal of Artificial Intelligence Research 67: 1-52

[9] Smith, M., & Kollock, P. (2019). Ethics of artificial intelligence and robotics. In The SAGE Handbook of Social Media Research Methods (pp. 441-454). Sage Publications Ltd.

[10] Obermeyer Z., Powers B., Vogeli C., Mullainathan S.(2019). Dissecting Racial Bias In An Algorithm Used To Manage The Health Of Populations Science 366(6464):447-453.

[11] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 77-91.

[12] Smith, C. (2018). Addressing bias in artificial intelligence in health care. JAMA, 320(23), 2407-2408.

[13] Smithsonian Magazine (2020). How Artificial Intelligence Can Be More Inclusive: Diverse Data Is Key To Avoiding Biased Algorithms [Online].

[14] Smithsonian Magazine (2021). How Can We Make Algorithms More Ethical?

[15] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453.

[16] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 77-91.

[17] Bender, M., Brownlowe-Banezziere, C., et al. (2021). The importance of diverse samples in research. Journal of Diversity Studies, 15(2), 45-62.

[18] Dwork C., Hardt M., Pitassi T., Reingold O., & Zemel R.S. (2012). Fairness through awareness. In Proceedings of Innovations in Theoretical Computer Science Conference (ITCS), 214-226.