NEURAL INFORMATION PROCESSING SYSTEMS

Beyond Protected Attributes: Disciplined Detection of Systematic Deviations in Data

Trustworthy & Socially Responsible Machine Learning @ NeurIPS 2022

Adebayo Oshingbesan
Research Engineer, IBM Research Africa

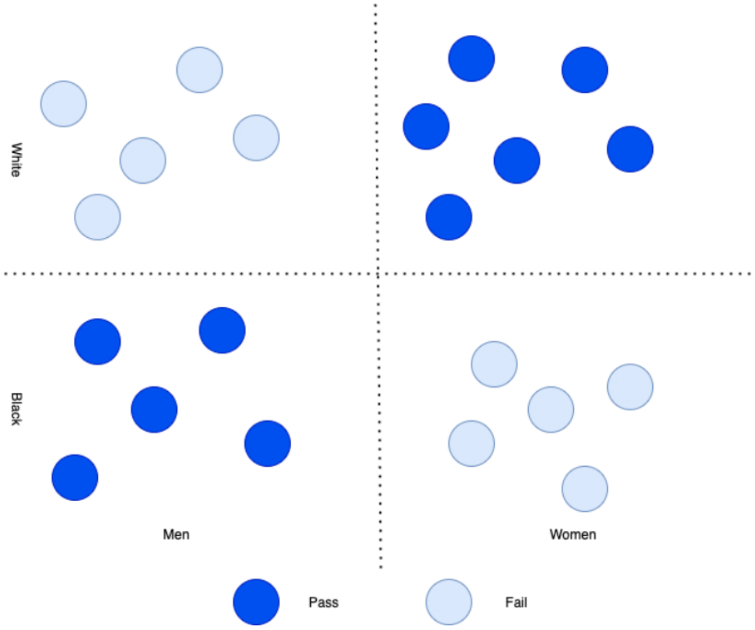
The Team: IBM Research Africa – AI Sciences






Systematic Deviations: Story So Far

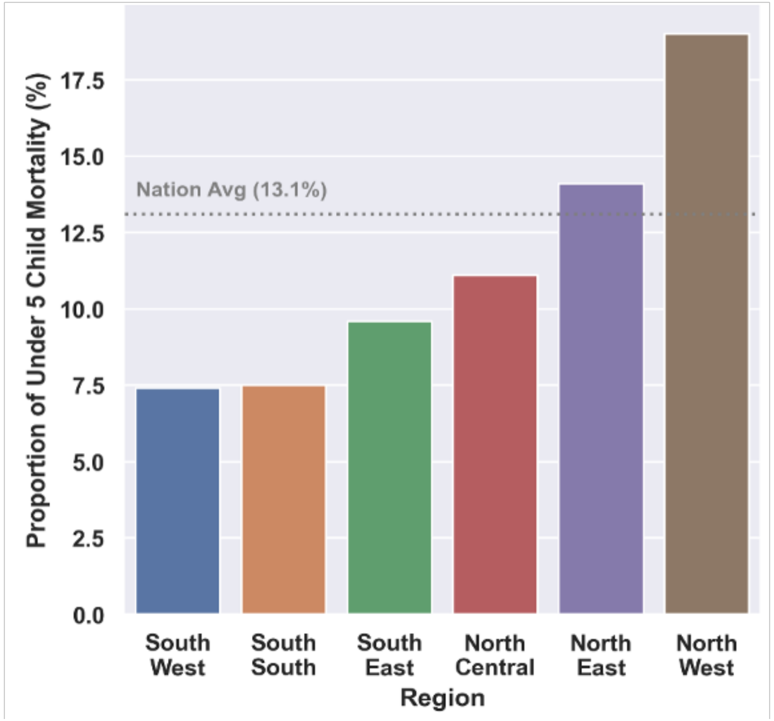
- Detecting systematic deviations helps in exploratory data analysis, data quality assessments, drifts detection, etc.
- Focus on group and individual level analysis on pre-selected attributes.
- Group /individual level analysis could be misleading (Kearns et al., 2018; Foulds et al., 2020).



Source: Kearns et al., 2018

Are Pre-selected Features Truly Enough?

- What coverage of features should our analysis have?
- How could we help domain experts when analysing data for systematic deviations?



Subgroup Discovery as a Solution

- Subgroup discovery is an association mining technique that finds interesting patterns in transactional databases that can be extended for tabular datasets.

- Previous works on subgroup discovery include techniques such as Apriori (Limmerich et al., 2014), Slice Finder (Chung et al., 2019), and FP-Growth (Pastor et al., 2021).

- These techniques are not scalable and require either the size or the extremity of the deviation to be set.

- We present Automatic Stratification (AutoStrat) - an efficient algorithm for automatically discovering interesting subgroups.
    - We validate with several datasets for different use cases and compare with other subgroup discovery techniques.

AutoStrat

- AutoStrat finds the anomalous subgroups with higher-than-average outcomes as compared to the global mean, p_i.

  – H_0: odds(y_i) = $\frac{p_i}{1-p_i}$, is constant for all subgroups.

  – H_1: odds(y_i) = $q * \frac{p_i}{1-p_i}$, where q, the odds multiplier, > 1, for some subgroups.

- We search for the subgroups with the most evidence of $q > 1$ by maximizing the Bernoulli likelihood ratios between these hypothesis .

$$\max_{q>1} \log \prod_{i \in S} \frac{\text{Bernoulli}\left(\frac{qp_i}{1-p_i+qp_i}\right)}{\text{Bernoulli}(p_i)} = \max_{q>1} \sum_{i \in S} y_i \cdot \ln(q) - \log(1 - p_i + q \cdot p_i)$$

- If the search space is limited to pre-selected features, the subgroup discovered is called a protected subgroup (PS). Otherwise, the subgroup is called beyond-protected subgroup (BPS).

Experimental Setup

- Datasets: Compas, Credit Card Client, OULAD Education Data

Dataset	Number of Records	Protected Attributes	Possible Subgroups (Without logical ORs)	Target	Outcome Proportion
Compas	4,743	Sex, Race	250,047 (432)	v_decile_score >5	0.2043
Credit Card	30,000	Sex, Education, Marriage	2.79E+62 (3.06E+21)	default payment next month	0.2212
OULAD	32,593	Gender, Disability	3.12E+13 (218,400)	final results = pass or distinction	0.3151

- Baselines: Beam search, Apriori (Limmerich et al., 2014), FP-Growth (Chung et al., 2019), Slice Finder (Pastor et al., 2021).

- Metrics: Lift, Support, Odds Ratio *(OR)*, Weighted Relative Accuracy (ϕ), Bernoulli Likelihood Statistic (Γ), *p-value, and runtime.*

Result: Comparison between Protected Subgroups & Non-Protected Subgroups Across Three Datasets

Dataset	Type	Subgroup	p-value	OR	$\Gamma(S)$
Compas	BPS	age_cat = Less than 25	0.0099	**9.33**	**274**
	PS	sex = Male AND race = African-American OR Native American	0.0099	1.86	70
Credit Card	BPS	PAY_0 = 2 OR 3 OR 4	0.0099	**11.55**	**1583**
	PS	MARRIAGE = 1 OR 2 OR 3 AND SEX = 1 AND EDUCATION = 2 OR 3	0.5445	1.27	40
OULAD	BPS	studied_credits = 90.0 - 655.0 AND region = NOT (IRELAND or WALES) AND imb_band = 0%-90%	0.0099	**2.26**	**309**
	PS	disability = Y	0.0099	1.42	43

Result (Contd.): Comparison of AutoStrat with Other Recently Proposed Algorithms for OULAD Dataset

Method	Subgroup	Lift	Support	OR	$\Gamma(S)$	$\phi(S)$	Time
AutoStrat (Ours)	imd_band=0% - 90% AND studied_credits = 90 - 120 OR 120 - 655 AND region = NOT(Ireland OR Wales)	1.47	0.2	2.26	**309**	**0.03**	25.44
Beam search	studied_credits=90 - 120	1.29	0.2	1.67	118	0.02	**0.72**
Apriori	studied_credits=90 - 120	1.29	0.2	1.67	118	0.02	6.72
Fp-growth	num_prev_attempts=0 AND studied_credits=90 - 120	1.31	0.16	1.68	109	0.02	11.73
Slice Finder	region=North Western AND num_prev_attempts=2	1.85	1.91E-3	**3.05**	9	0	260.77

Conclusion and Future Work

- We described AutoStrat - an efficient algorithm for divergent subgroup discovery.

- One limitation of AutoStrat, like other subgroup discovery algorithms, is the need to bin continuous features. Future works include supporting continuous variables directly.

- Also, while we only focused on the most divergent subgroup in this paper, we would be extending the analysis to multiple returned subgroups in future works.

Thank you! Asante!






Paper

Code

Email: adebayo.oshingbesan1@ibm.com

References

- Michael Kearns, Seth Neel, Aaron Roth, and Zhiwei Steven Wu. Preventing fairness gerrymandering: Auditing and learning for subgroup fairness. In International Conference on Machine Learning, pages 2564–2572. PMLR, 2018
- James R Foulds, Rashidul Islam, Kamrun Naher Keya, and Shimei Pan. An intersectional definition of fairness. In 2020 IEEE 36th International Conference on Data Engineering (ICDE), pages 1918–1921. IEEE, 2020
- Daniel B Neill and Tarun Kumar. Fast multidimensional subset scan for outbreak detection and characterization. Online Journal of Public Health Informatics, 5(1), 2013
- Zhe Zhang and Daniel B Neill. Identifying significant predictive bias in classifiers. arXiv preprint arXiv:1611.08292, 2016
- Joshua D Habiger and Edsel A Pena. Randomised p-values and nonparametric procedures in multiple testing. Journal of nonparametric statistics, 23(3):583–604, 2011
- Bernard V North, David Curtis, and Pak C Sham. A note on the calculation of empirical p values from monte carlo procedures. The American Journal of Human Genetics, 71(2):439–441, 2002
- Daniel B Neill. Fast subset scan for spatial pattern detection. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 74(2):337–360, 2012
- Florian Lemmerich. Novel techniques for efficient and effective subgroup discovery. Bayerische Julius-Maximilians-Universitaet Wuerzburg (Germany), 2014.
- Yeounoh Chung, Tim Kraska, Neoklis Polyzotis, Ki Hyun Tae, and Steven Euijong Whang. Slice finder: Automated data slicing for model validation. In 2019 IEEE 35th International Conference on Data Engineering (ICDE), pages 1550–1553. IEEE, 2019.
- Eliana Pastor, Luca de Alfaro, and Elena Baralis. Looking for trouble: Analyzing classifier behavior via pattern divergence. In Proceedings of the 2021 International Conference on Management of Data, pages 1400–1412, 2021