Semi-supervised Feature Selection for Efficient Detection of Systemic Deviations to Develop Trustworthy AI

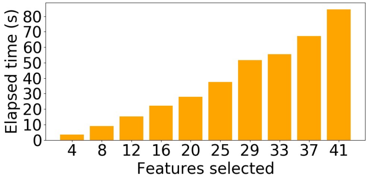
IBM Research | Africa

Girmaw Abebe Tadesse, William Ogallo, Aisha Walcott-Bryant, Skyler Speakman
IBM Research | Africa

Email: girmaw.abebe.tadesse@ibm.com

1. Introduction

- Identifying systemic deviations in datasets and model outputs helps to achieve trustworthy AI
- Multiple techniques have been proposed in the state-of-the-art[1] to detect systemic deviations in tabular data, but the computational complexity grows with the dimension of the feature space.

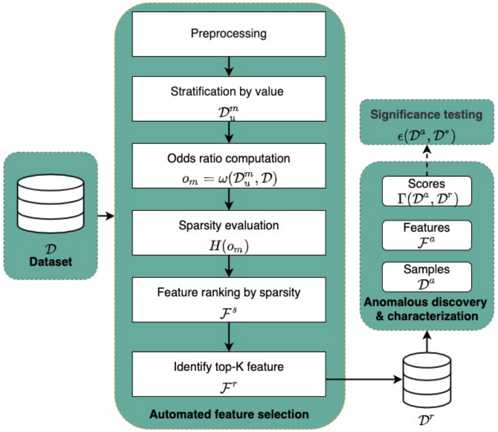


- We propose a **model-free** and sparsity-based automated feature selection (**SAFS**) framework for efficient discovery of anomalous patterns

2. Proposed framework

- SAFS employs sparsity of the odds ratios of feature values
- Deviation across values of a given feature is encoded using Hoyer-based sparsity metric[2]
- Top K features are selected based on sparsity-based ranking and fed into the MDSS-based subgroup discovery technique[1]

$$H(o_m) = \left(\sqrt{C_m} - \frac{\sum_{u=1}^{C_m} o_m^u}{\sum_{u=1}^{C_m} o_m^{2\,u}}\right)\left(\sqrt{C_m} - 1\right)^{-1}$$



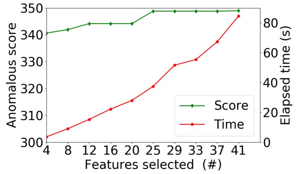
3. Experiments

- **Dataset**: Medical Information Mart for Intensive Care (MIMIC-III)[3]
- Outcome: Patients death within 28 days of the onset of ICU admission
- N=19,658 patients, and M = 41 features

- **Baseline methods**
- Mutual-information based Filter[4]
- Backward-elimination based Wrapper[5], and
- Embedded techniques (XGBoost, CatBoost and Committee)[6]

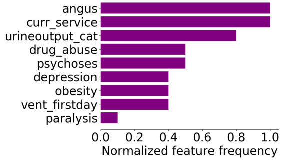
- **Performance metric**
- Computation time and characterization of identified subgroup

Results



Top K	Size	Odds Ratio (95% CI)	P-Value
4	4312	2.48 (2.3, 2.67)	< 0.01
8	2811	3.00 (2.75, 3.26)	< 0.01
12	4383	2.48 (2.31, 2.67)	< 0.01
16	4383	2.48 (2.31, 2.67)	< 0.01
20	4383	2.48 (2.31, 2.67)	< 0.01
25	3078	2.93 (2.7, 3.18)	< 0.01
29	3078	2.93 (2.7, 3.18)	< 0.01
33	3078	2.93 (2.7, 3.18)	< 0.01
37	3078	2.93 (2.7, 3.18)	< 0.01
41	4218	2.54 (2.36, 2.73)	< 0.01

- **3x** reduction in computation time using just **half** of the features
- Consistent subgroup detection
- Superior to the baselines



	Baseline methods					**Proposed**
			Embedded methods			
K	Filter	Wrapper	XGB	CatB	Committee	SAFS
4	337.19	311.27	337.19	339.23	337.19	**340.61**
8	340.61	311.27	337.19	340.61	340.61	**341.90**
12	340.61	311.27	340.61	340.61	340.61	**344.13**
16	340.61	316.09	340.61	**344.13**	340.61	**344.13**
20	340.61	340.86	340.61	**344.13**	**344.13**	**344.13**
25	340.61	346.64	348.53	348.45	348.53	**348.77**
29	344.59	346.64	**348.77**	**348.77**	**348.77**	**348.77**
33	**348.77**	346.64	**348.77**	**348.77**	**348.77**	**348.77**
37	348.45	**348.91**	**348.91**	**348.91**	**348.91**	348.77
41	**348.91**	**348.91**	**348.91**	**348.91**	**348.91**	**348.91**

4. Conclusions

- We proposed a sparsity-based automated feature selection (SAFS) framework for anomalous subgroup discovery **without** training a particular model.
- SAFS outperformed multiple baseline feature selection methods and achieved more than **3×** reduction in computational time but with competitive detection performance using just **half** of the features.
- Future work aims to extend SAFS to select layers and nodes, e.g., for adversarial detection in deep learning frameworks

References

1. Neill et al. Fast subset scan for multivariate event detection. Statistics in Medicine (2013)
2. Hurley and Rickard. Comparing measures of sparsity. IEEE Tran. on Info. Theory 55, 10 (2009)
3. Johnson et al., MIMIC-III, a freely accessible critical care database.
4. Estévez et al. Normalized mutual information feature selection. IEEE Tran. on Neu. Net. (2009)
5. Miao and Niu. A survey on feature selection. Pro. Comp. Sci. 91 (2016), 919–926.
6. Wanjiru et al. Automated Supervised Feature Selection for Differentiated Patterns of Care. (2021) arXiv preprint arXiv:2111.03495