IEEE SMC 2023
Oahu, Hawaii

香 港 大 學
THE UNIVERSITY OF HONG KONG

華東師範大學
EAST CHINA NORMAL UNIVERSITY

Meta Pattern Concern Score: A Novel Evaluation Measure with Human Values for Multi-classifiers

Yanyun Wang [1], Dehui Du [2] and Yuanhao Liu [2]

[1] Department of Computer Science, The University of Hong Kong, Hong Kong, China

[2] Software Engineering Institute, East China Normal University, Shanghai, China

Presenter: Yanyun Wang

Presentation Video, The 2023 IEEE International Conference on Systems, Man, and Cybernetics

Brief Introduction

- To date, while advanced classifiers have been increasingly used in real-world safety-critical applications, how to properly evaluate the black-box models given specific human values remains a concern in the community.

- Such human values include punishing error cases of different severity in varying degrees and making compromises in general performance to reduce specific dangerous cases.

- In this paper, we propose a novel evaluation measure named Meta Pattern Concern Score based on the abstract representation of probabilistic prediction and the adjustable threshold for the concession in prediction confidence, to introduce these human values into multi-classifiers.

Background

$$\mathcal{M} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1c} \\ m_{21} & m_{22} & \cdots & m_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ m_{c1} & m_{c2} & \cdots & m_{cc} \end{bmatrix}$$

$$TP = m_{rr}$$

$$FN = \sum_{j=1}^{c} m_{rj} - m_{rr}$$

$$FP = \sum_{i=1}^{c} m_{ir} - m_{rr}$$

$$TN = \sum_{i=1}^{c}\sum_{j=1}^{c} m_{ij} - \sum_{i=1}^{c} m_{ir} - \sum_{j=1}^{c} m_{rj} + m_{rr}$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F_1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

.....

$$\omega^* = arg\,min\,\mathcal{L}(\omega) = arg\,min\,\frac{1}{|\mathcal{X}|}\sum_{x \in \mathcal{X}} l(x, \omega)$$

$$\mathcal{L}_{MS} = -\frac{1}{2|\mathcal{X}|}\sum_{i=1}^{|\mathcal{X}|}\sum_{j=1}^{c}(\hat{y}_i^{(j)} - y_i^{(j)})^2$$

$$\mathcal{L}_{CE} = -\frac{1}{|\mathcal{X}|}\sum_{i=1}^{|\mathcal{X}|}\sum_{j=1}^{c} y_i^{(j)}\, log(\hat{y}_i^{(j)})$$

.....

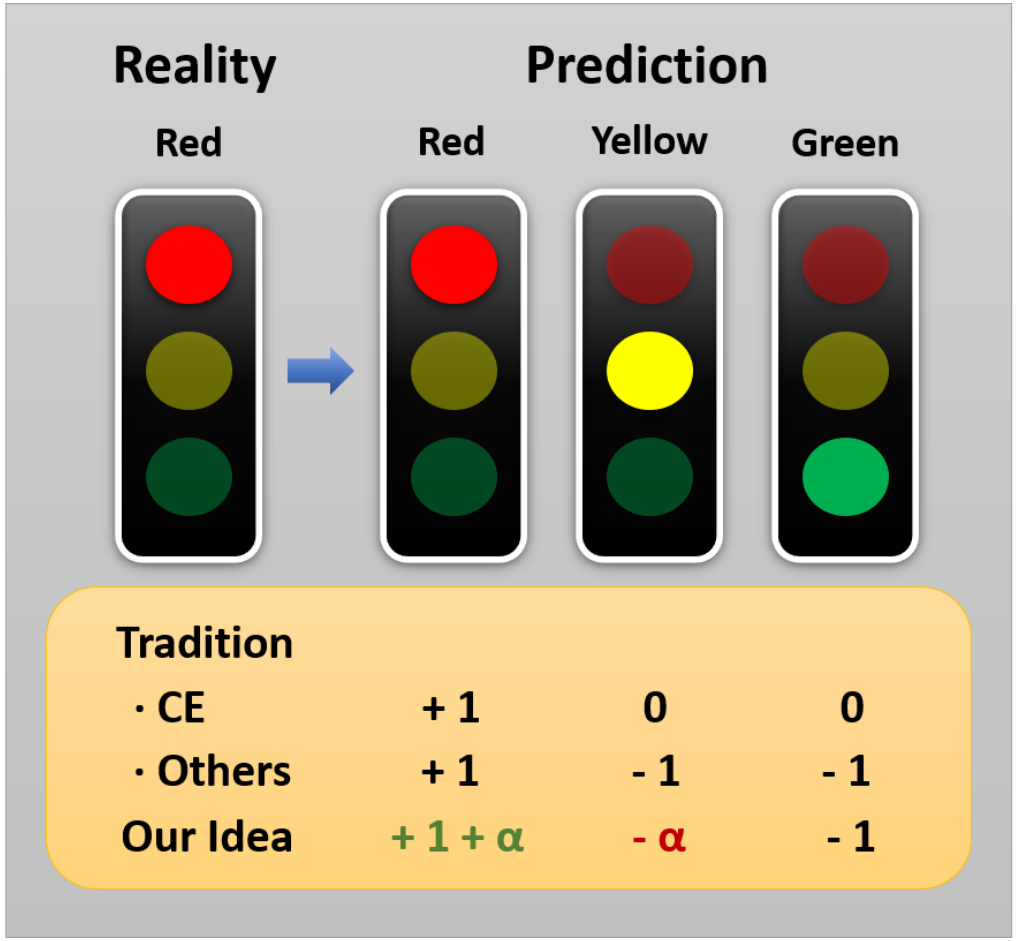
Motivation

- Just uniformly assessing different models built for different tasks is gradually found to be limited and unsatisfying. Because beyond the general performance, there are usually some specific human values to be concerned with and weighed in such areas.

- Many specific cases in our human society are not just black-or-white, so instead of equally punishing all kinds of incorrect predictions with different destructiveness, we allow assigning specific weights individually for every single error case.

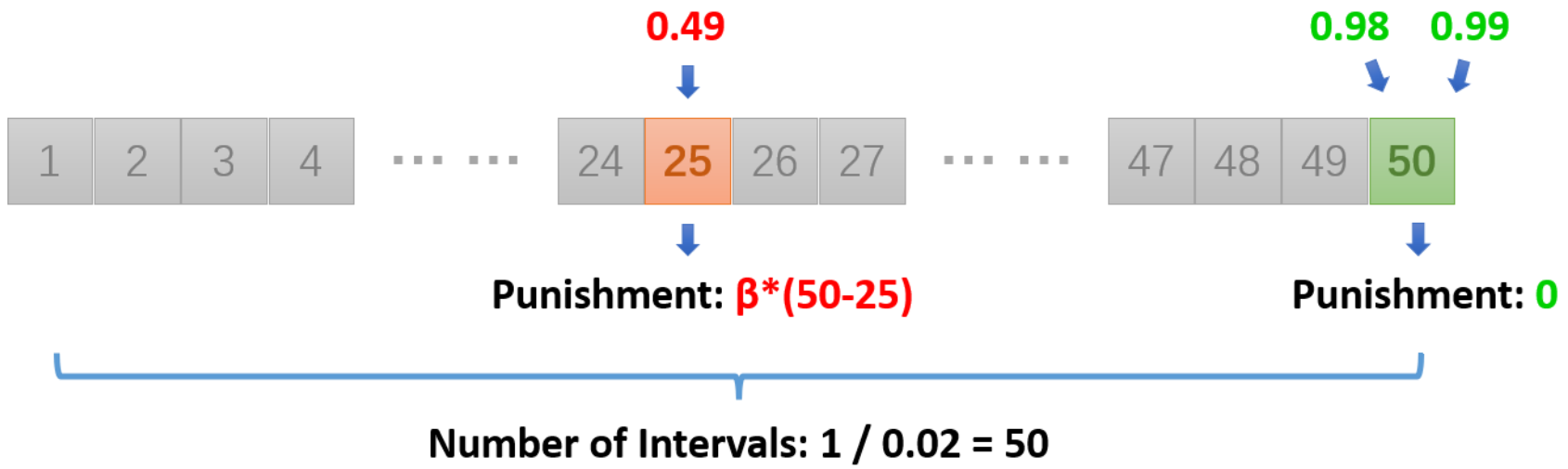
- In order to satisfy some safety-critical requirements in practice, we would rather make certain compromises in overall performance to reduce specific dangerous cases.

Considering Human Values in MPCS

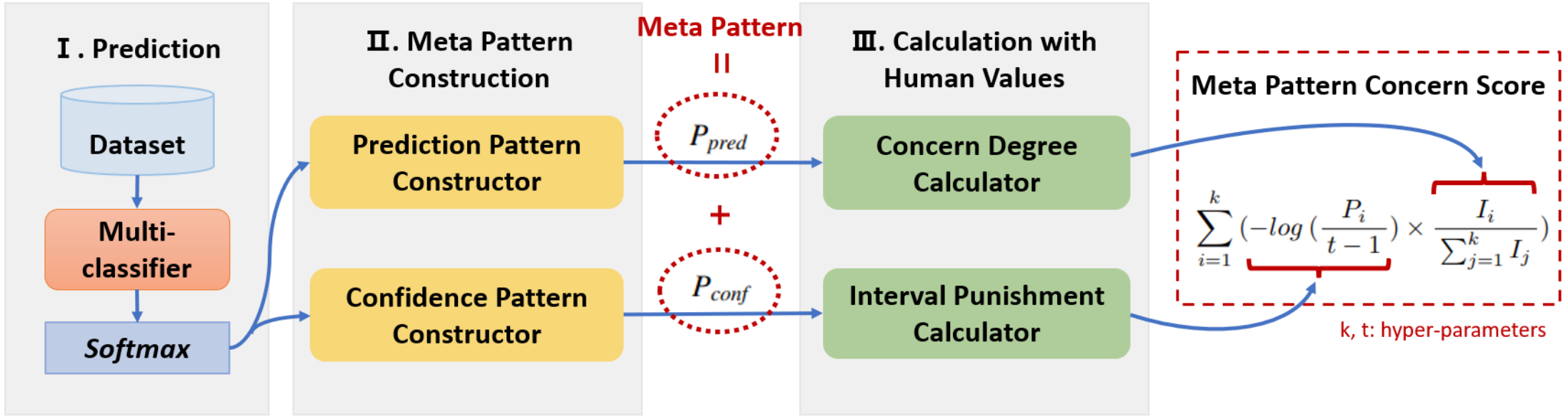


Considering Human Values in MPCS

Training Size	Classifier	Prediction Result	Number of Sample	Prediction Confidence	Cross Entropy Loss
100	A	Correct	99	0.99	0.741920
		Wrong	1	0.49	
	B	Correct	100	0.98	0.877392
		Wrong	0	-	



Detailed Design of MPCS



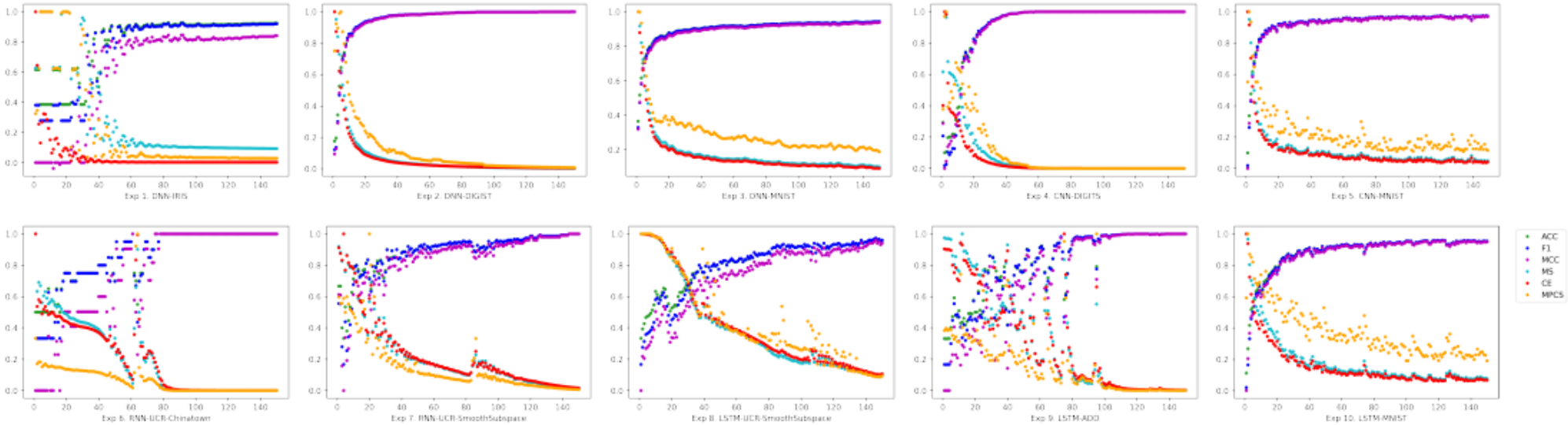
$$P_{pred} = [l_1, l_2, \ldots, l_k]$$

$$P_{conf} = [C(l_1), C(l_2), \ldots, C(l_k)] \quad \text{where} \quad C(l_i) = \begin{cases} \lfloor t \cdot c_i \rfloor, & l_i = l \\ t - \lfloor t \cdot c_i \rfloor - 1, & l_i \neq l \end{cases}$$

$$\mathcal{I} = [D(l_1), D(l_2), \ldots, D(l_k)] \quad \text{where} \quad D(l_i) = \begin{cases} \sum_{l_j \in P_{pred},\, j \neq i} D(l_j), & l_i = l \\ f_{\mathcal{R}}, & l_i \neq l,\ [l, li] \in r_\alpha \\ 1, & l_i \neq l,\ [l, li] \notin r_\alpha \end{cases}$$

Evaluation – Effectiveness as An Evaluation Measure

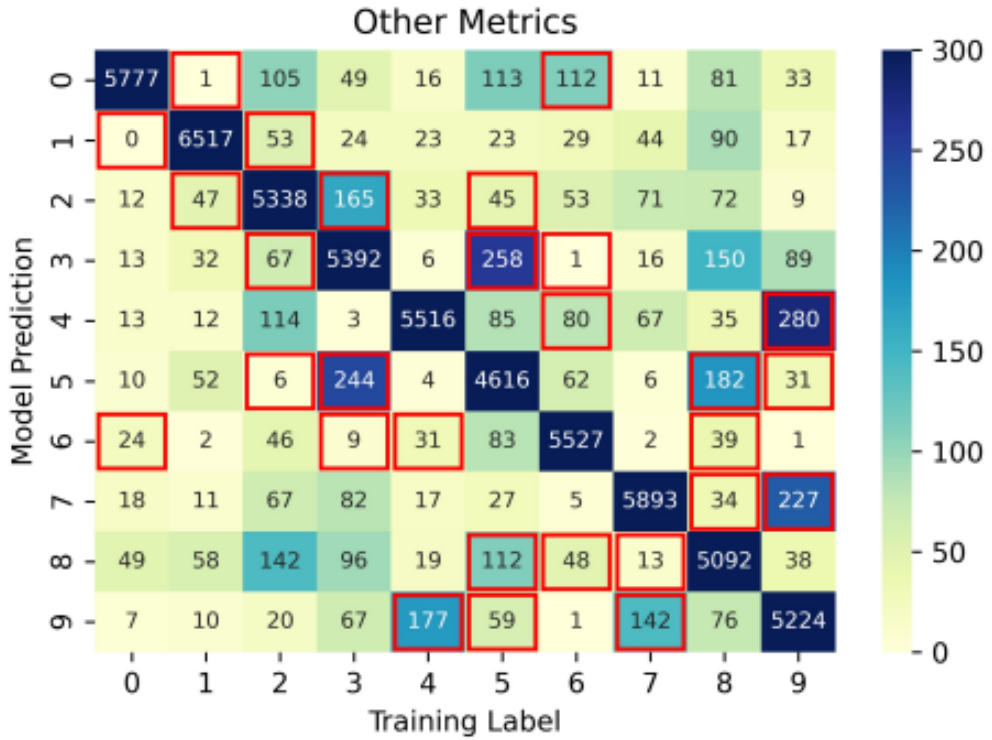
ID	Model	Dataset	Spearman Similarity				
			ACC	F1	MCC	MS	CE
Exp. 1	MLP	IRIS	-0.9091	-0.9092	-0.9011	**0.9873**	**0.9827**
Exp. 2		DIGITS	**-0.9886**	**-0.9936**	**-0.9935**	**0.9989**	**0.9989**
Exp. 3		MNIST	**-0.9881**	**-0.9886**	**-0.9880**	**0.9880**	**0.9841**
Exp. 4	CNN	DIGITS	-0.9106	-0.9106	-0.9107	0.9547	0.9547
Exp. 5		MNIST	-0.9231	-0.9229	-0.9220	0.9299	0.9168
Exp. 6	RNN	UCR-CT	-0.9576	-0.9549	-0.9587	0.9618	0.9619
Exp. 7		UCR-SS	**-0.9846**	**-0.9844**	**-0.9827**	**0.9929**	**0.9928**
Exp. 8	LSTM	UCR-SS	-0.9382	-0.9352	-0.9373	**0.9700**	**0.9718**
Exp. 9		ADD	**-0.9721**	**-0.9735**	**-0.9706**	**0.9815**	**0.9862**
Exp. 10		MNIST	**-0.9757**	**-0.9754**	**-0.9760**	**0.9774**	**0.9763**

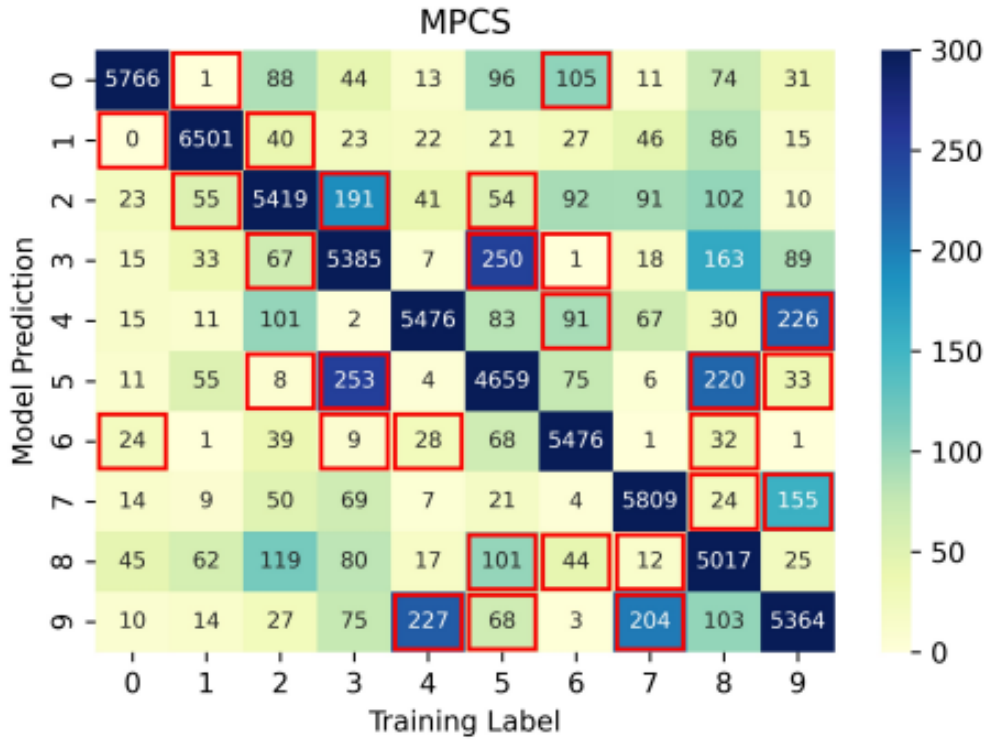


Evaluation – Computational Efficiency

ID	Time Cost ($\times 10^{-3}s$)					
	ACC	F1	MCC	MS	CE	MPCS
Exp. 1	0.28	1.76	2.01	0.49	0.49	4.27
Exp. 2	0.33	1.62	2.79	0.59	0.49	13.66
Exp. 3	26.15	50.21	88.63	33.88	33.74	497.49
Exp. 4	31.56	32.98	33.90	31.80	35.55	**43.98**
Exp. 5	692.03	696.62	702.07	693.26	788.27	**710.92**
Exp. 6	7.25	8.34	8.17	7.26	7.59	**7.08**
Exp. 7	21.30	22.27	22.34	21.39	22.43	**22.40**
Exp. 8	38.92	40.05	39.99	39.04	40.07	**38.89**
Exp. 9	228.87	230.86	232.93	229.05	245.07	**230.59**
Exp. 10	2672.93	2703.11	2746.43	2685.93	2927.46	**2802.63**

Evaluation – Ability to Introduce Human Values





Contribution

- For the first time, we provide a general idea to introduce the two kinds of specific human values into multi-classifiers.
- Different from common metrics having a fixed form all the time, the MPCS allows people to flexibly declare what they care more about the model in different practices, and try to cater to their specific will to pick out the optimal model, with the premise of not violating the common metrics too much and having a similar time cost as them.
- The MPCS is expected to support the customized evaluation and even training of multi-classifiers in real-world practice, especially in safety-critical areas with various human values to be considered.

Thank You!

Acknowledgement

This work is funded by the NSFC under Grant No.61972153 and by Yingfeng Capital (Shandong) Co., Ltd.

Contact Information

Dehui Du: dhdu@sei.ecnu.edu.cn

Yanyun Wang: yynwang@connect.hku.hk