

Visual Methods for Analyzing Probabilistic Classification Data

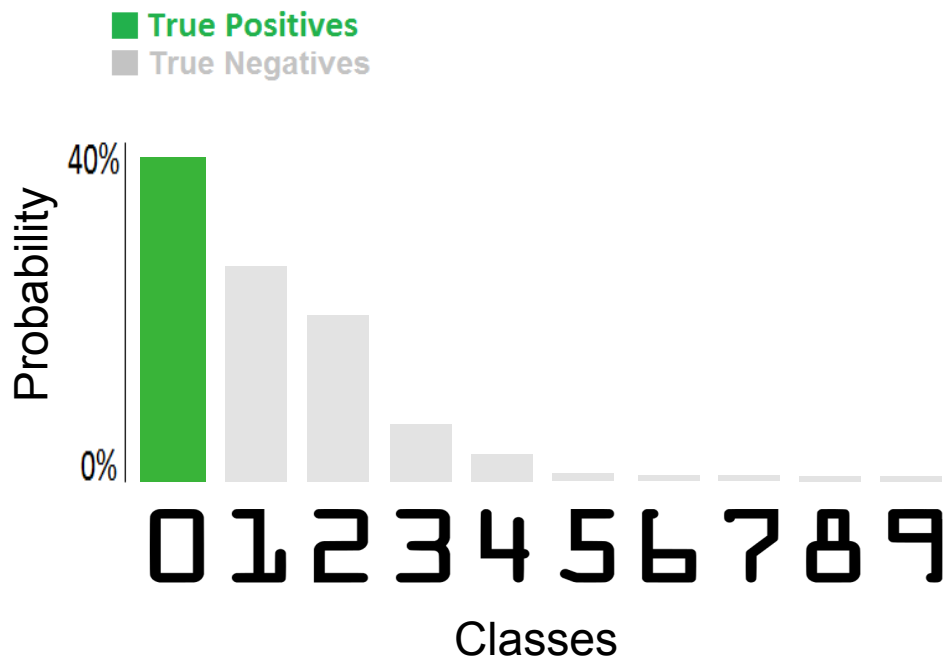
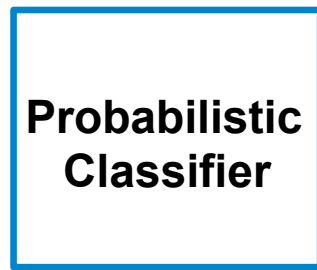
Bilal Alsallakh, Allan Hanbury, Silvia Miksch, Andreas Rauber, Helwig Hauser

Presented by Yingsai Dong
 Information Visualization
 UBC Computer Science
 Nov. 2014

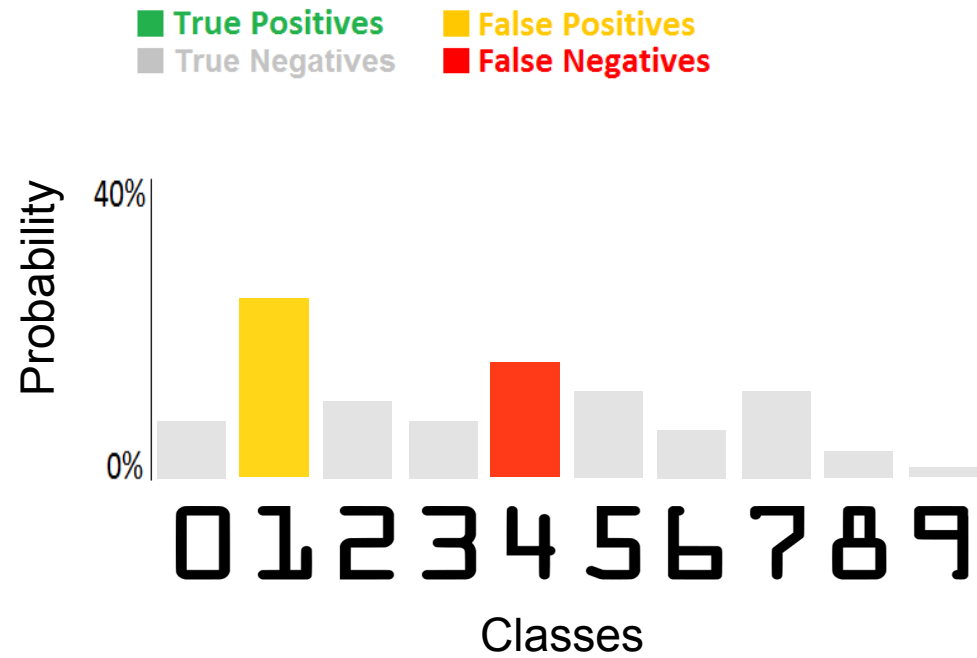
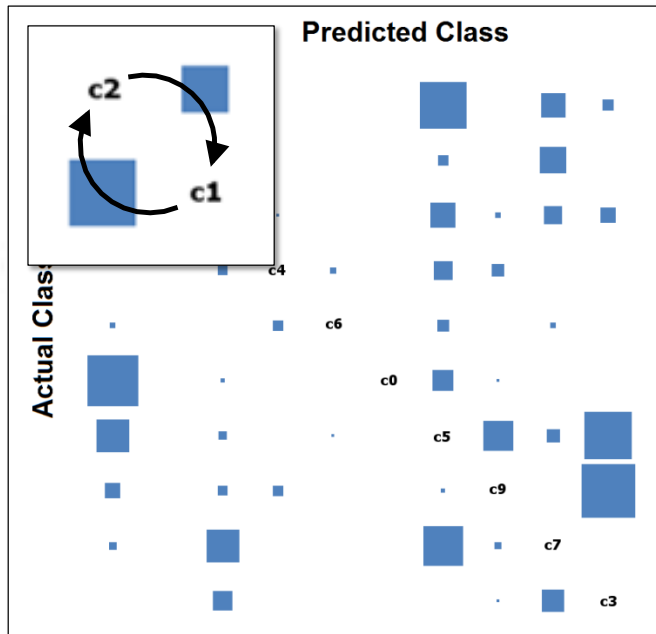
Probabilistic Multi-class Classification



Sample



Probabilistic Multi-class Classification



Analyzing Probabilistic Classification Data

Tasks

(A) Analyze probability distributions (of samples in a class)

Both overall distribution and sub-distributions by classification results

(B) Select samples

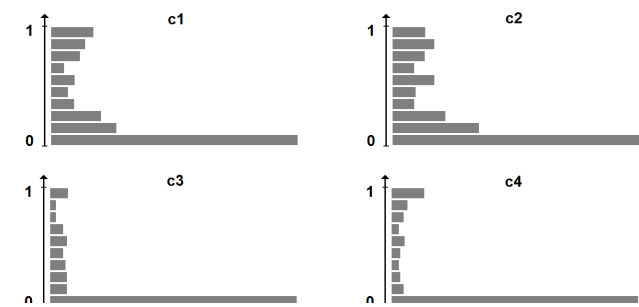
by probabilities, classification results, data features, certain class confusions, ...

(C) Analyze separability (of *correct* and *incorrect* classifications)

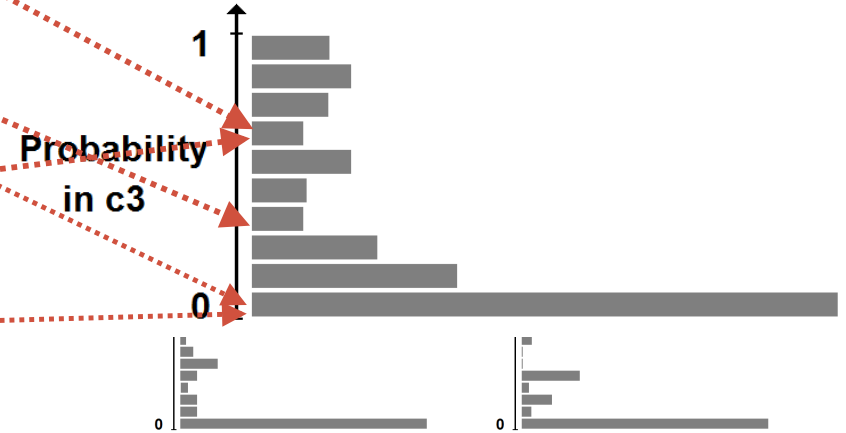
can certain **FPs** / **FNs** be separated from **TPs** / **TNs** by *data features*?

Task A: Analyze Probability Distributions

Data Features					Class Probabilities									
X1	Y1	...	Y8	Actual	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10
100,0	90,0	...	98,0	c8	0,38	0,00	0,00	0,00	0,31	0,31	0,00	0,00	0,00	0,00
2,0	2,0	...	6,0	c3	0,00	0,65	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
25,0	25,0	...	,0	c3	0,00	0,69	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,31
23,0	23,0	...	,0	c4	0,00	0,00	0,74	0,00	0,00	0,00	0,00	0,26	0,00	0,00
20,0	20,0	...	,0	c1	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,34
73,0	16,0	...	20,0	c6	0,00	0,00	0,00	0,73	0,00	0,27	0,00	0,00	0,00	0,00
49,0	49,0	...	,0	c4	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
:	:	...	:	:	:	:	:	:	:	:	:	:	:	:
22,0	22,0	...	24,0	c7	0,00	0,35	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
17,0	17,0	...	17,0	c3	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00
7,0	7,0	...	7,0	c3	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00



All Samples

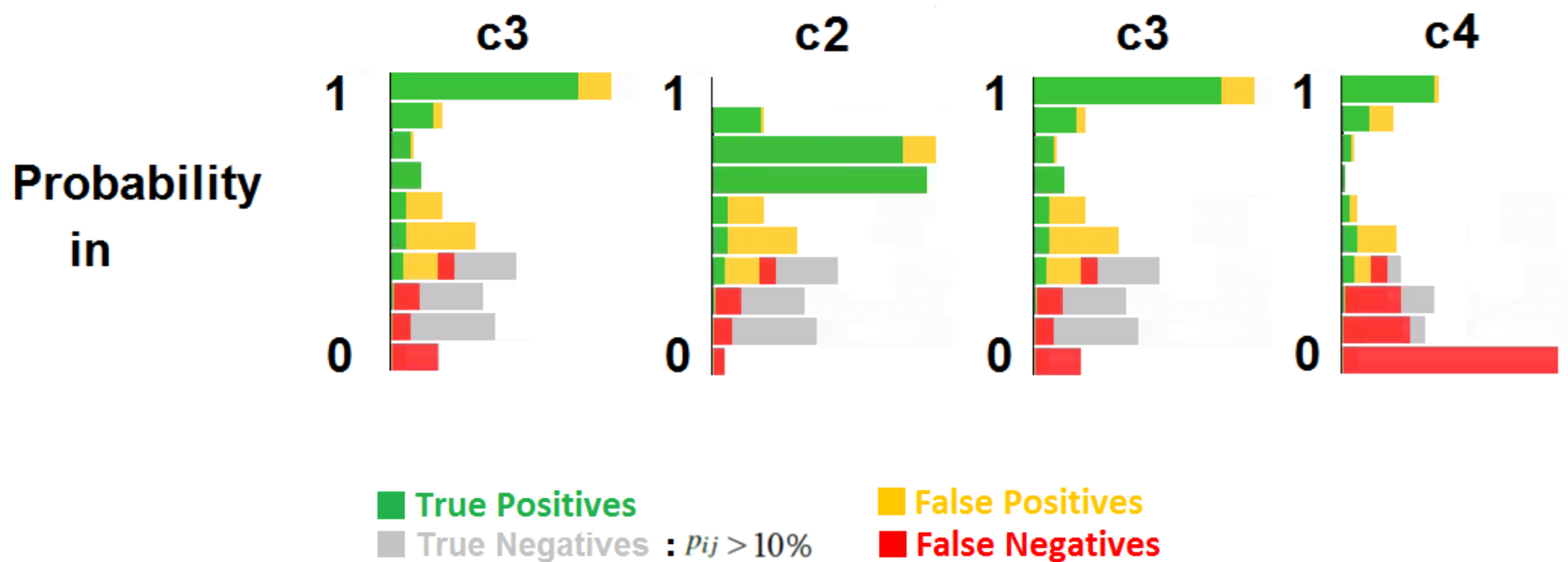


3
4

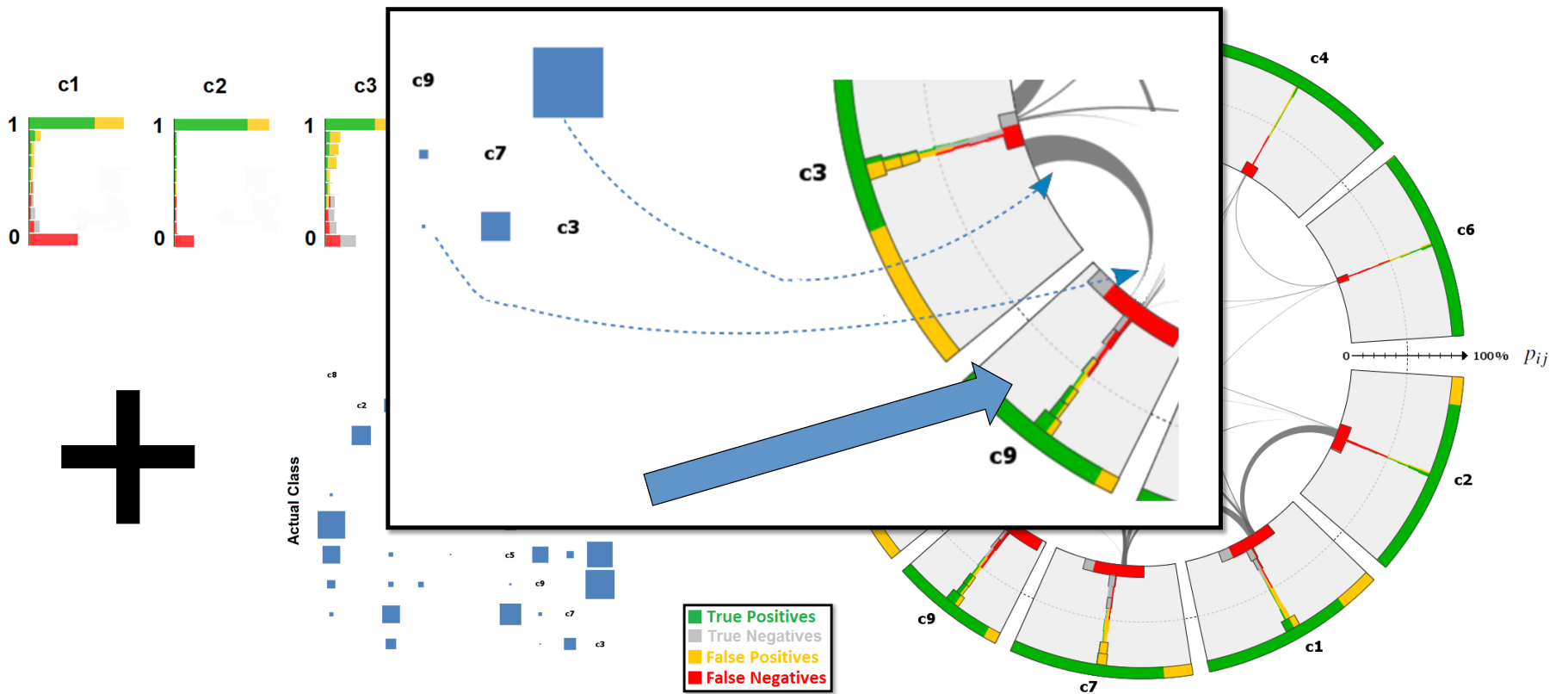
0,63
0,00
0,00

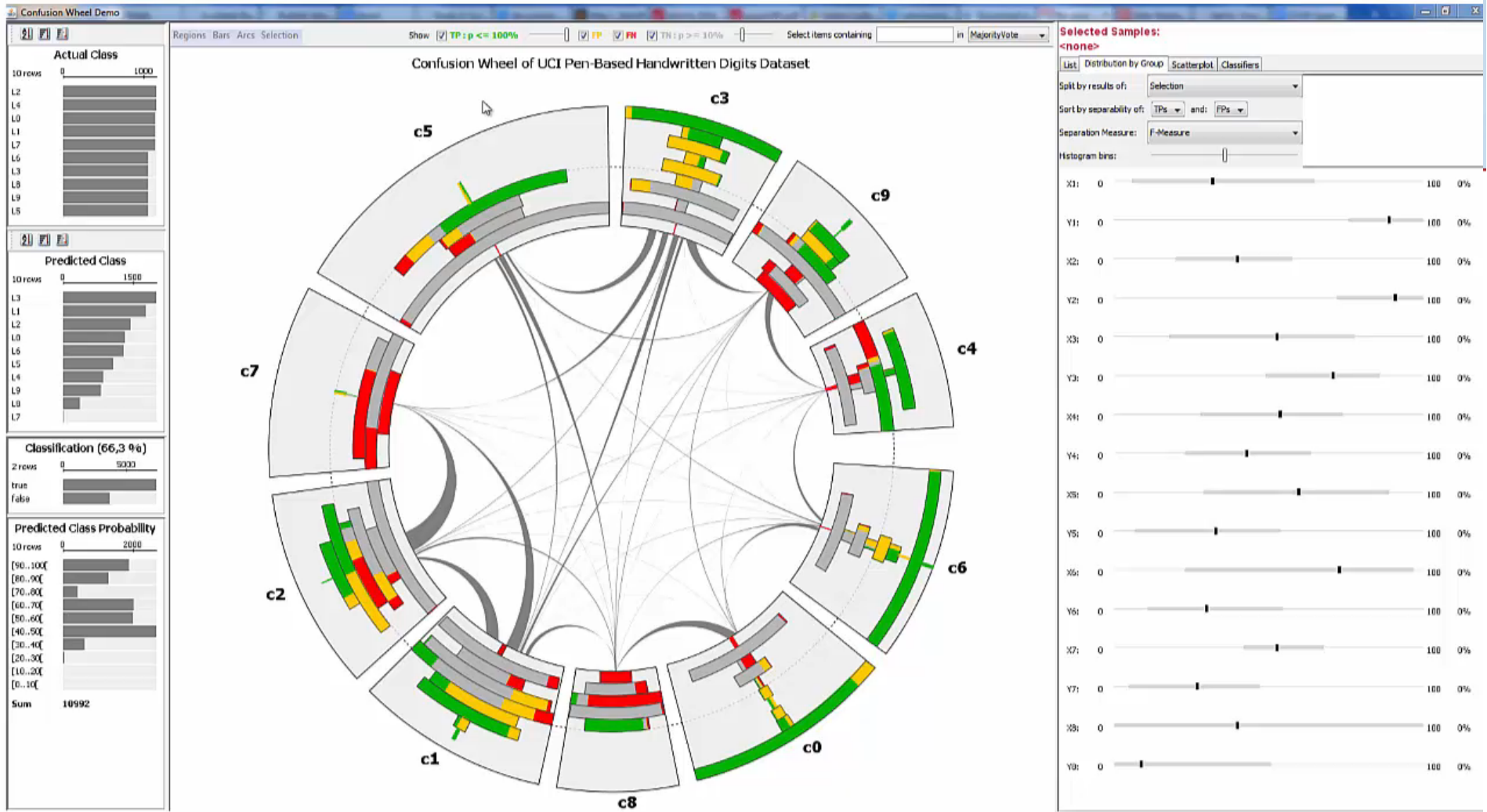
Probability
in c3

Probability Sub-Histograms



Augmenting Class Confusion





Task B: Interactive Selection

Data Features				
	X1	Y1	...	Y8
1	100,0	90,0	...	98,0
2	69,0	2,0	...	6,0
3	28,0	25,0	...	,0
4	74,0	23,0	...	,0
5	33,0	20,0	...	,0
6	73,0	16,0	...	20,0
7	65,0	49,0	...	,0
		⋮		⋮
8	53,0	22,0	...	24,0
9	32,0	,0	...	17,0
0	49,0	,0	...	7,0

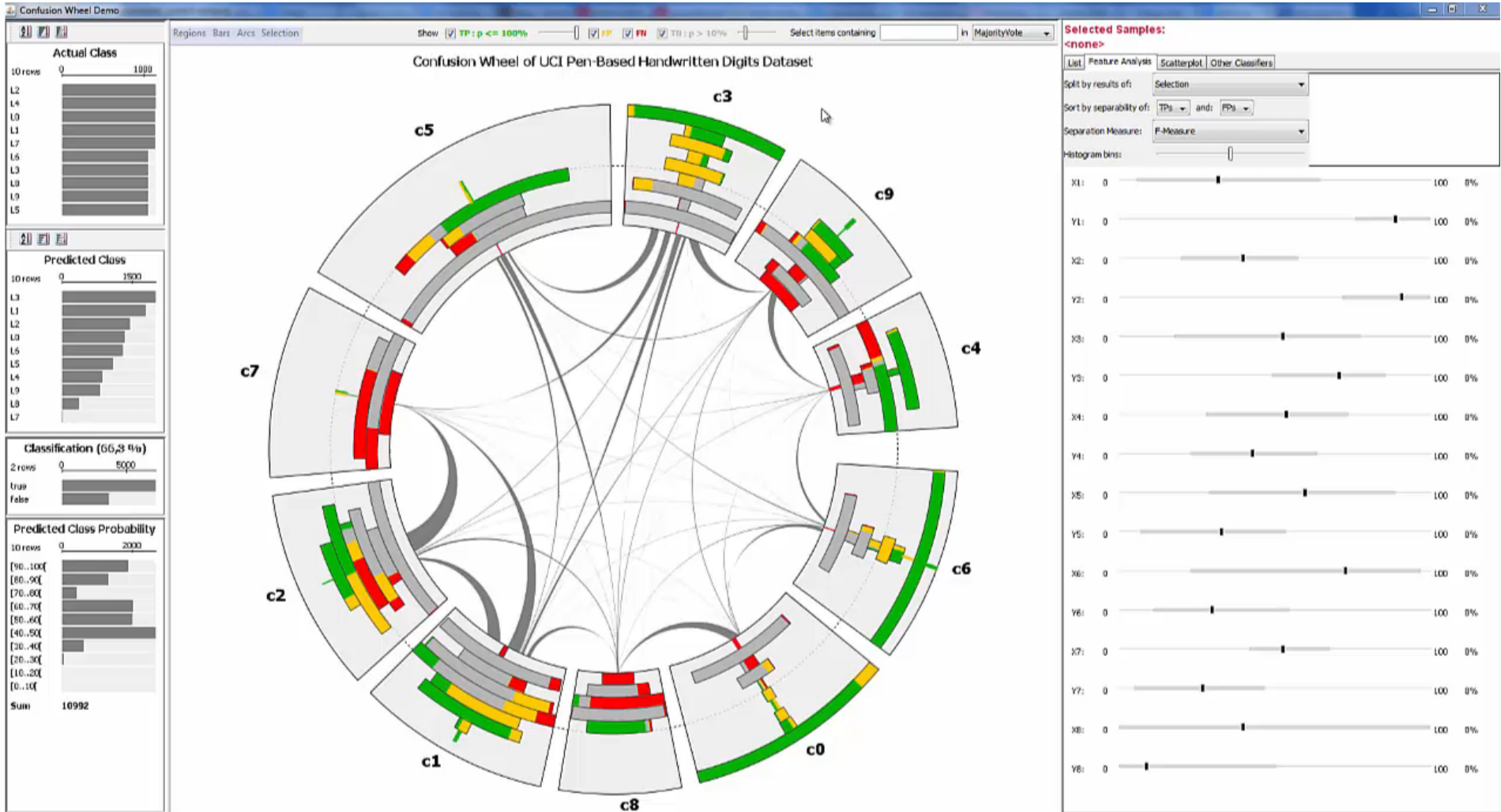


Selected Samples:
L2 ⇌ L7

List	Distribution by Group	Scatterplot	Classifiers			
MajorityVote	X1	Y1	X2	Y2	X3	
L2	0.0	98.0	36.0	100.0	80.0	
L2	0.0	93.0	62.0	100.0	100.0	
L2	0.0	97.0	42.0	100.0	67.0	
L2	0.0	100.0	67.0	100.0	86.0	
L2	24.0	100.0	100.0	99.0	93.0	
L2	21.0	92.0	85.0	100.0	90.0	
L2	0.0	97.0	41.0	100.0	58.0	
L2	0.0	87.0	55.0	100.0	100.0	
L2	0.0	96.0	40.0	100.0	60.0	
L2	0.0	93.0	45.0	100.0	74.0	
L2	0.0	80.0	30.0	100.0	74.0	
L2	0.0	84.0	40.0	100.0	76.0	
L2	0.0	100.0	65.0	99.0	70.0	
L2	0.0	82.0	47.0	100.0	75.0	
L2	0.0	100.0	67.0	94.0	83.0	
L2	0.0	99.0	48.0	100.0	100.0	
L2	0.0	92.0	39.0	100.0	67.0	
L2	0.0	98.0	40.0	100.0	78.0	
L2	0.0	84.0	25.0	100.0	58.0	

555 items

Classified with k -NN (w. $k = 5$)



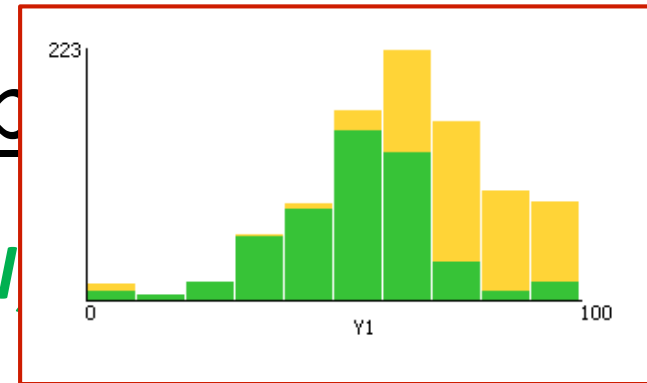
Task C: Group Sep

Can a **data feature** separate *correctly* classified samples?

Sort features by separability:

- *p*-Value

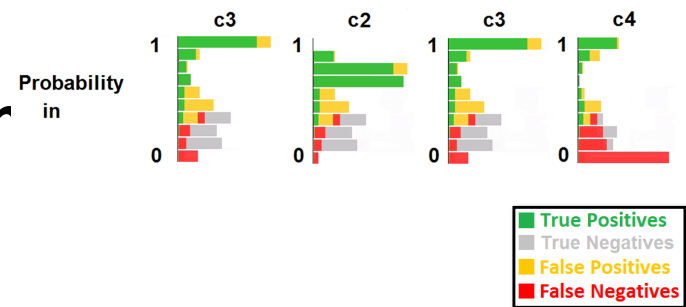
Data Features			
X1	Y1	...	Y8
100,0	98,0		98,0
69,0	2,0		6,0
28,0	25,0		,0
74,0	23,0		,0
33,0	20,0		,0
73,0	16,0	...	20,0
65,0	49,0		,0
	⋮		⋮
53,0	22,0		24,0
32,0	,0	...	17,0
49,0	,0		7,0



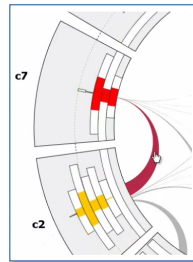
■ True Positives
■ False Positives

Tasks

(A) Analyze probability distribution



(B) Select samples

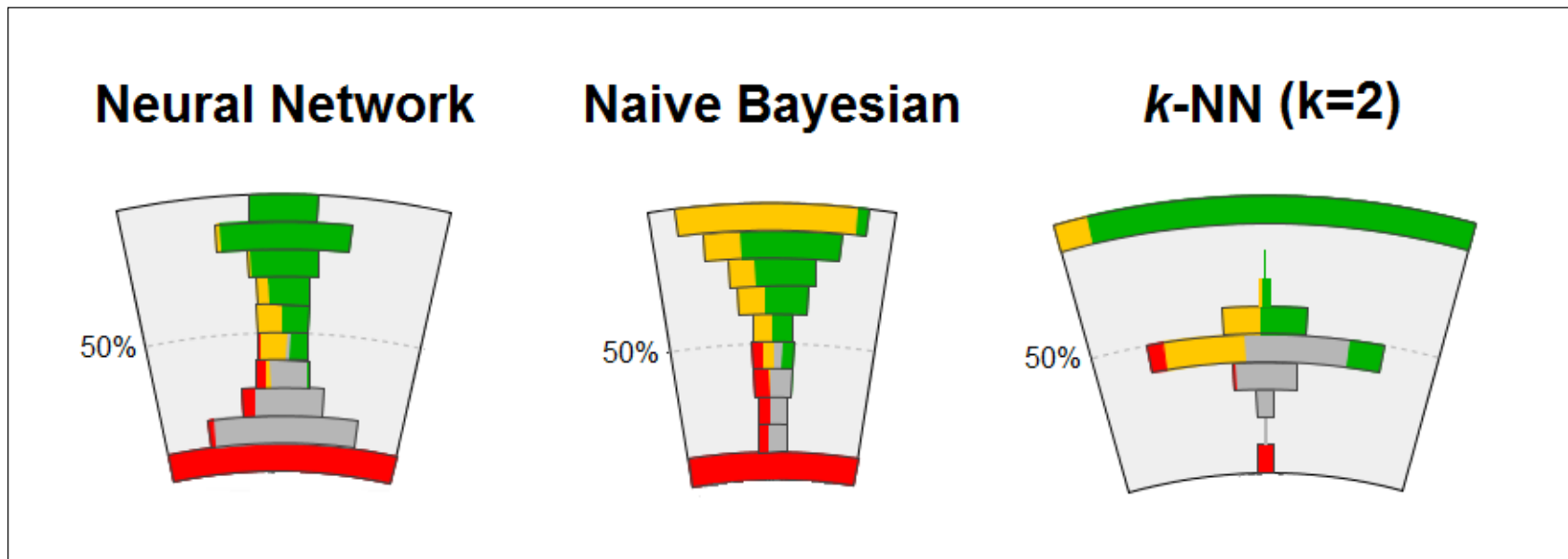


(C) Analyze separability



Usage Scenarios

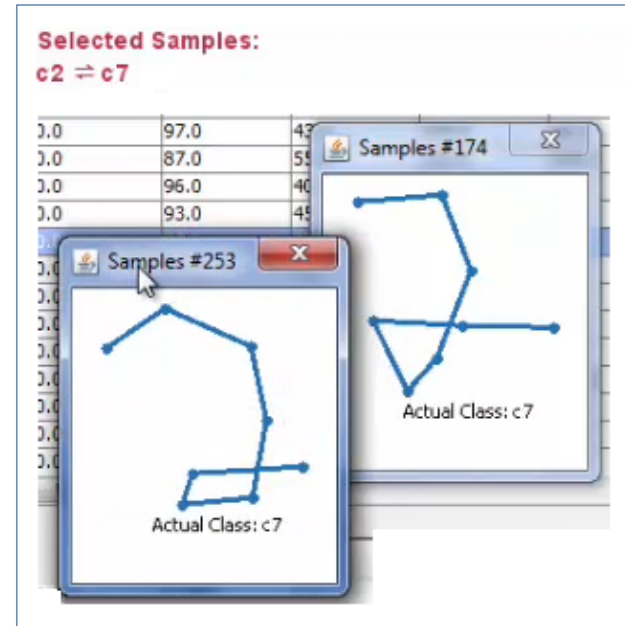
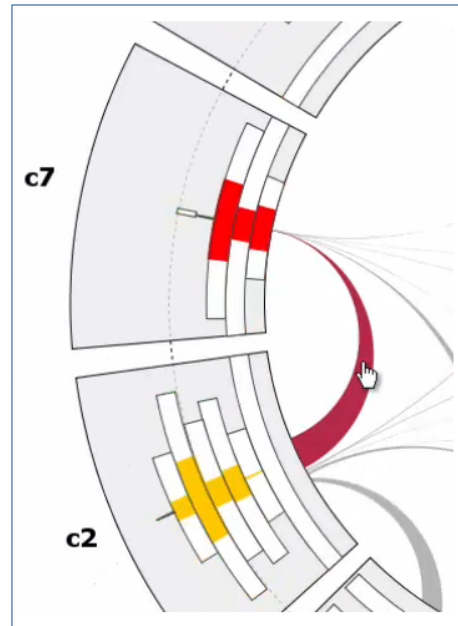
Analyzing Classifier Behavior



Based on Task A (analyzing probability distribution)

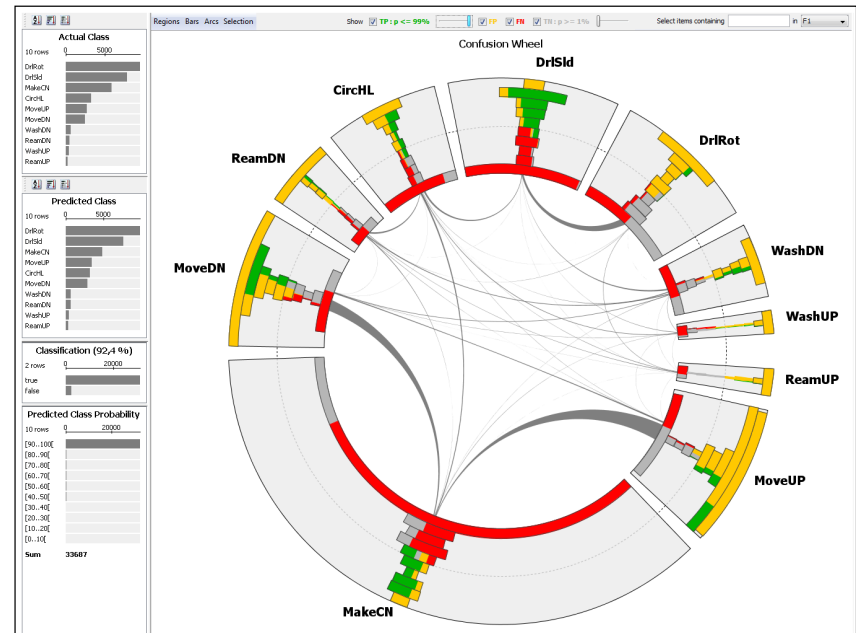
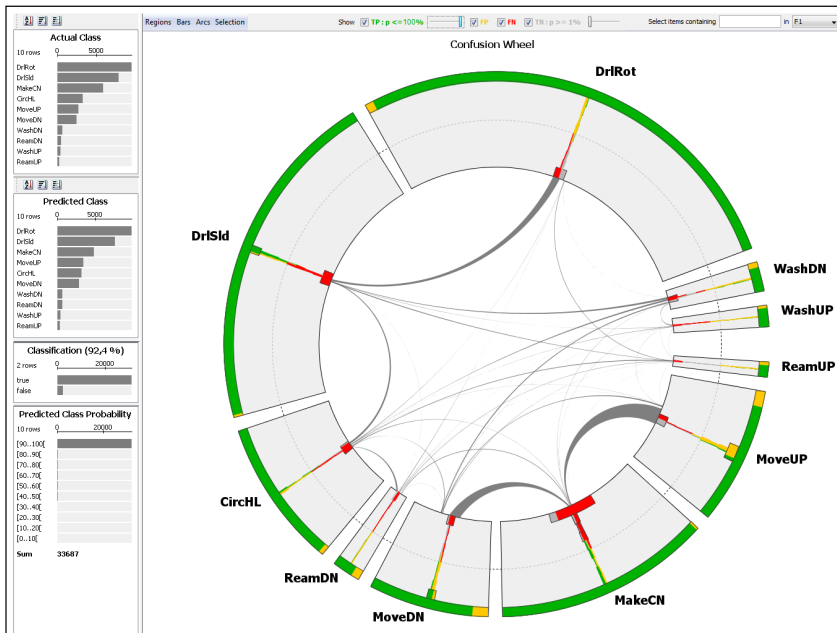
Visual Inspection

- False Positives
- False Negatives



Based on Task B (interactive sample selection)

Visual Inspection



TPs > 90% are filtered

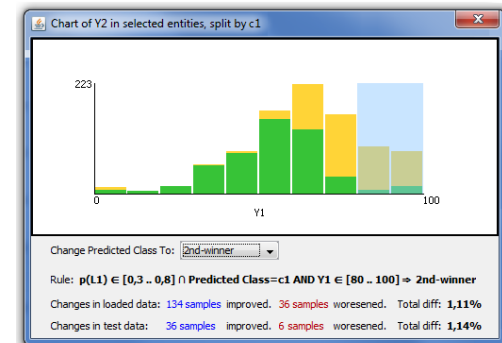
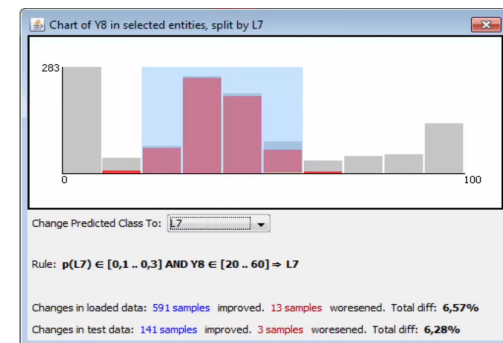
Defining Post-Classification Rules

Rectify certain **false negatives**

$$\mathbf{R}_{\text{FN}_j} : q_1(s_i) \wedge \dots \wedge q_k(s_i) \wedge l^P(s_i) \neq c_j \Rightarrow l^P(s_i) \leftarrow c_j$$

Rectify certain **false positives**

$$\mathbf{R}_{\text{FP}_j} : q_1(s_i) \wedge \dots \wedge q_k(s_i) \wedge l^P(s_i) = c_j \Rightarrow l^P(s_i) \leftarrow c_{j'} : r(s_i, c_{j'}) = 2$$



Based on Task C (separability analysis)

Limitations

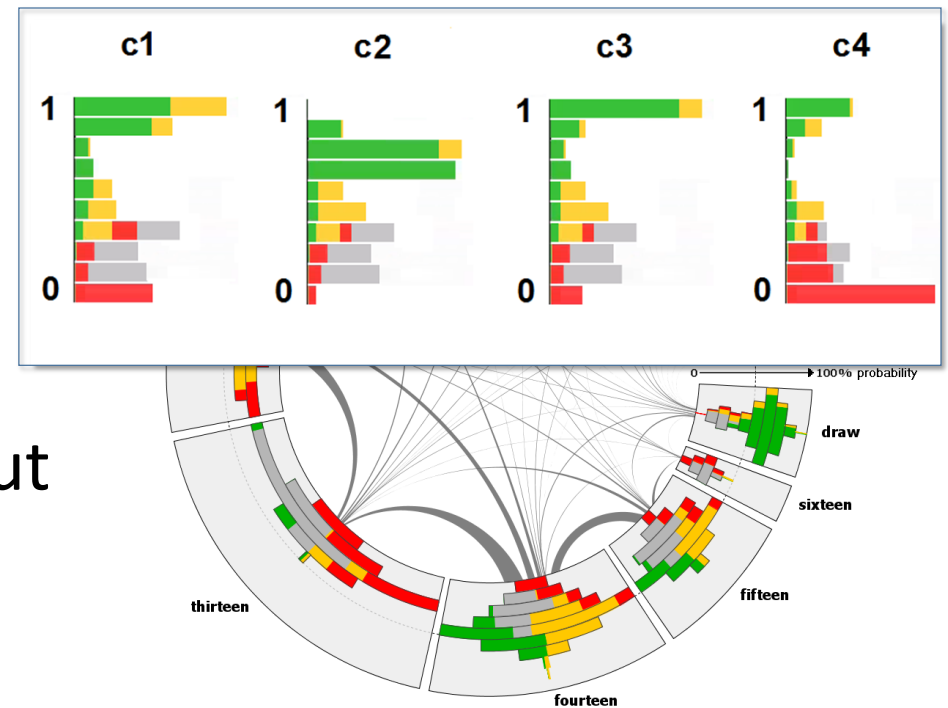
Visual Complexity

- Too much information
- Needs extensive training

Scalability in # of classes

Issues with the radial layout

- Arcs instead of straight bars
- Histograms with no base lines



Conclusion

Classification probabilities

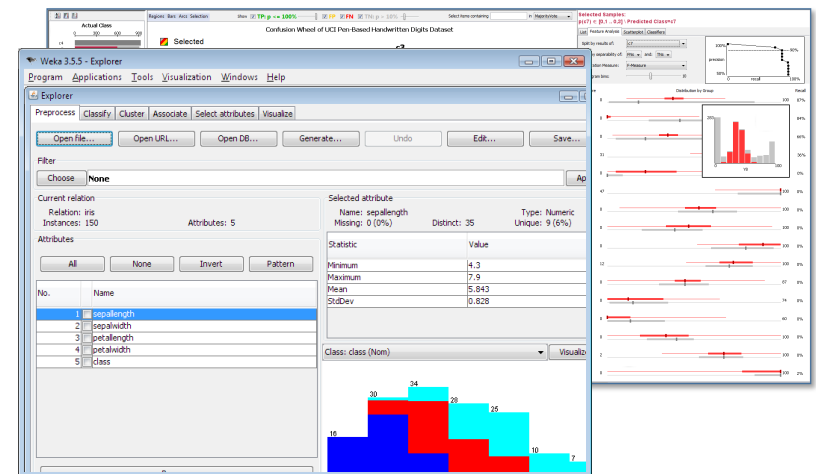
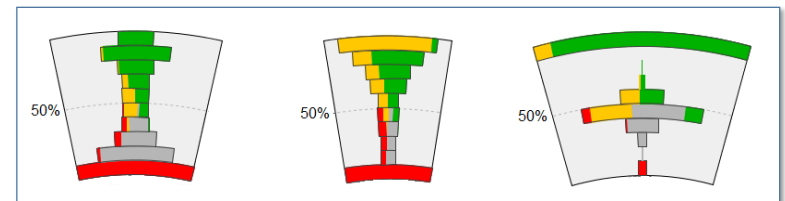
Rich of information

Explain classifier behavior

Interactive exploration

Reveals several insights

Guides new improvements



<http://www.cvast.tuwien.ac.at/ConfusionAnalysis>

Summary

- What?
 - High-dimensional probabilistic classification data
- Why?
 - Task: Analyze probability distribution, select samples, analyze separability
- How?
 - Stacked histogram (bar), wheel layout, boxplot, multiple-views.
- Scale
 - # Classes up to 20. # samples up to tens of thousands.

Thanks!