National Research University Higher School of Economics
Faculty of Computer Science
School of Data Analysis and Artificial Intelligence

Explainable analytics: understanding causes, correcting errors, and achieving increasingly perfect accuracy from the nature of distinguishable patterns

Presenters: M. Kirdin, K. Ostudina

Authors: Hao-Ting Pai, Chung-Chian Hsu

Published in: *Scientific Reports* volume 12 (2022)

Moscow, 2024

Outline

- Problem Statement
- Methods
- Experiments & Results
- Conclusion
- References

Problem Statement

The article addresses the challenge of making classification and prediction errors **explainable**. These errors are caused by issues like faulty or inconsistent data, mismatches between data types and analytical methods, as well as the increasing complexity of large datasets, commonly referred to as the **curse of dimensionality**.

Methods

Methods

The authors propose **Transparent Classification method.**

Requirements:

- An algorithm that can visualize causes of prediction errors in a network and therefore make them traceable and correctable.

As is:

- It uses pure positive and negative patterns to classify data

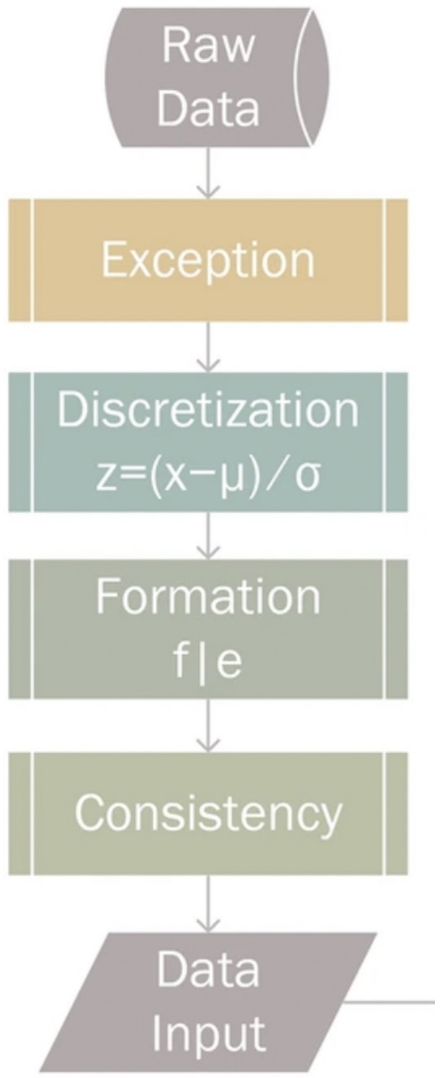
Data Preprocessing

Raw Data

Instance				Class
C0	C1	C2	C3	
a1	b1	1.1	0.1	0
a1	b1	1.0	0.1	1
a1	b1	0.9	1.9	0
a1	b1		0.2	0
a1	b2	1.0	0.1	0
a1	b2	9.7	1.8	1
a2	b1	9.9	1.9	1
a2	b2	9.8	1.8	1
a2	b1	1.1	0.1	0
a1	b2	9.6	0.2	0
a1	b2	1.0	1.9	0
a1	b1		1.8	0
a2	b2	1.2	0.2	0
a2	b1	9.9	0.1	0
a2	b1	8.9	1.7	0
a2	b2	0.9	1.8	1
a2	b2	9.5	1.9	1
a2	b2	9.6	1.8	1

Data Pre-processing

<table>
<tr><th>Exception</th><th colspan="2">Discretization</th><th colspan="4">Formation</th><th colspan="5">Consistency</th><th colspan="5">Data Input</th></tr>
<tr><th>C2</th><th>C2</th><th>C3</th><th>C0</th><th>C1</th><th>C2</th><th>C3</th><th>C0</th><th>C1</th><th>C2</th><th>C3</th><th>Class</th><th>C0</th><th>C1</th><th>C2</th><th>C3</th><th>Class</th></tr>
<tr><td>1.1</td><td>(1.1-5.3)/4.3=-1</td><td>-1</td><td>0|a1</td><td>1|b1</td><td>2|-1</td><td>3|-1</td><td>~~0~~</td><td>~~2~~</td><td>~~4~~</td><td>~~7~~</td><td>~~0~~</td><td>0</td><td>2</td><td>4</td><td>8</td><td>0</td></tr>
<tr><td>1.0</td><td>-1</td><td>-1</td><td>0|a1</td><td>1|b1</td><td>2|-1</td><td>3|-1</td><td>~~0~~</td><td>~~2~~</td><td>~~4~~</td><td>~~7~~</td><td>~~1~~</td><td>0</td><td>2</td><td>6</td><td>7</td><td>0</td></tr>
<tr><td>0.9</td><td>-1</td><td>1</td><td>0|a1</td><td>1|b1</td><td>2|-1</td><td>3|1</td><td>0</td><td>2</td><td>4</td><td>8</td><td>0</td><td>0</td><td>3</td><td>4</td><td>7</td><td>0</td></tr>
<tr><td>NA</td><td>NA</td><td>-1</td><td>0|a1</td><td>1|b1</td><td>2|NA</td><td>3|-1</td><td>0</td><td>2</td><td>6</td><td>7</td><td>0</td><td>0</td><td>3</td><td>5</td><td>8</td><td>1</td></tr>
<tr><td>1.0</td><td>-1</td><td>-1</td><td>0|a1</td><td>1|b2</td><td>2|-1</td><td>3|-1</td><td>0</td><td>3</td><td>4</td><td>7</td><td>0</td><td>1</td><td>3</td><td>5</td><td>8</td><td>1</td></tr>
<tr><td>9.7</td><td>1</td><td>1</td><td>0|a1</td><td>1|b2</td><td>2|1</td><td>3|1</td><td>0</td><td>3</td><td>5</td><td>8</td><td>1</td><td>1</td><td>2</td><td>4</td><td>7</td><td>0</td></tr>
<tr><td>9.9</td><td>1</td><td>1</td><td>0|a2</td><td>1|b1</td><td>2|1</td><td>3|1</td><td>~~1~~</td><td>~~2~~</td><td>~~5~~</td><td>~~8~~</td><td>~~1~~</td><td>0</td><td>3</td><td>5</td><td>7</td><td>0</td></tr>
<tr><td>9.8</td><td>1</td><td>1</td><td>0|a2</td><td>1|b2</td><td>2|1</td><td>3|1</td><td>1</td><td>3</td><td>5</td><td>8</td><td>1</td><td>0</td><td>3</td><td>4</td><td>8</td><td>0</td></tr>
<tr><td>1.1</td><td>-1</td><td>-1</td><td>0|a2</td><td>1|b1</td><td>2|-1</td><td>3|-1</td><td>1</td><td>2</td><td>4</td><td>7</td><td>0</td><td>0</td><td>2</td><td>6</td><td>8</td><td>0</td></tr>
<tr><td>9.6</td><td>1</td><td>-1</td><td>0|a1</td><td>1|b2</td><td>2|1</td><td>3|-1</td><td>0</td><td>3</td><td>5</td><td>7</td><td>0</td><td>1</td><td>3</td><td>4</td><td>7</td><td>0</td></tr>
<tr><td>1.0</td><td>-1</td><td>1</td><td>0|a1</td><td>1|b2</td><td>2|-1</td><td>3|1</td><td>0</td><td>3</td><td>4</td><td>8</td><td>0</td><td>1</td><td>2</td><td>5</td><td>7</td><td>0</td></tr>
<tr><td>NA</td><td>NA</td><td>1</td><td>0|a1</td><td>1|b1</td><td>2|NA</td><td>3|1</td><td>0</td><td>2</td><td>6</td><td>8</td><td>0</td><td>1</td><td>3</td><td>4</td><td>8</td><td>1</td></tr>
<tr><td>1.2</td><td>-1</td><td>-1</td><td>0|a2</td><td>1|b2</td><td>2|-1</td><td>3|-1</td><td>1</td><td>3</td><td>4</td><td>7</td><td>0</td><td>1</td><td>3</td><td>5</td><td>8</td><td>1</td></tr>
<tr><td>9.9</td><td>1</td><td>-1</td><td>0|a2</td><td>1|b1</td><td>2|1</td><td>3|-1</td><td>1</td><td>2</td><td>5</td><td>7</td><td>0</td><td>1</td><td>3</td><td>5</td><td>8</td><td>1</td></tr>
<tr><td>8.9</td><td>1</td><td>1</td><td>0|a2</td><td>1|b1</td><td>2|1</td><td>3|1</td><td>~~1~~</td><td>~~2~~</td><td>~~5~~</td><td>~~8~~</td><td>~~0~~</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>0.9</td><td>-1</td><td>1</td><td>0|a2</td><td>1|b2</td><td>2|-1</td><td>3|1</td><td>1</td><td>3</td><td>4</td><td>8</td><td>1</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>9.5</td><td>1</td><td>1</td><td>0|a2</td><td>1|b2</td><td>2|1</td><td>3|1</td><td>1</td><td>3</td><td>5</td><td>8</td><td>1</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>9.6</td><td>1</td><td>1</td><td>0|a2</td><td>1|b2</td><td>2|1</td><td>3|1</td><td>1</td><td>3</td><td>5</td><td>8</td><td>1</td><td></td><td></td><td></td><td></td><td></td></tr>
</table>



Identifying Distinguishable Patterns

Positive Observations ($\boldsymbol{PO}$): These are instances in the dataset that belong to the positive class.

Positive Patterns ($\boldsymbol{pp_{\alpha-\beta}}$): A set of features that consistently appears in the positive observations. To find these patterns, the algorithm intersects different positive observations.

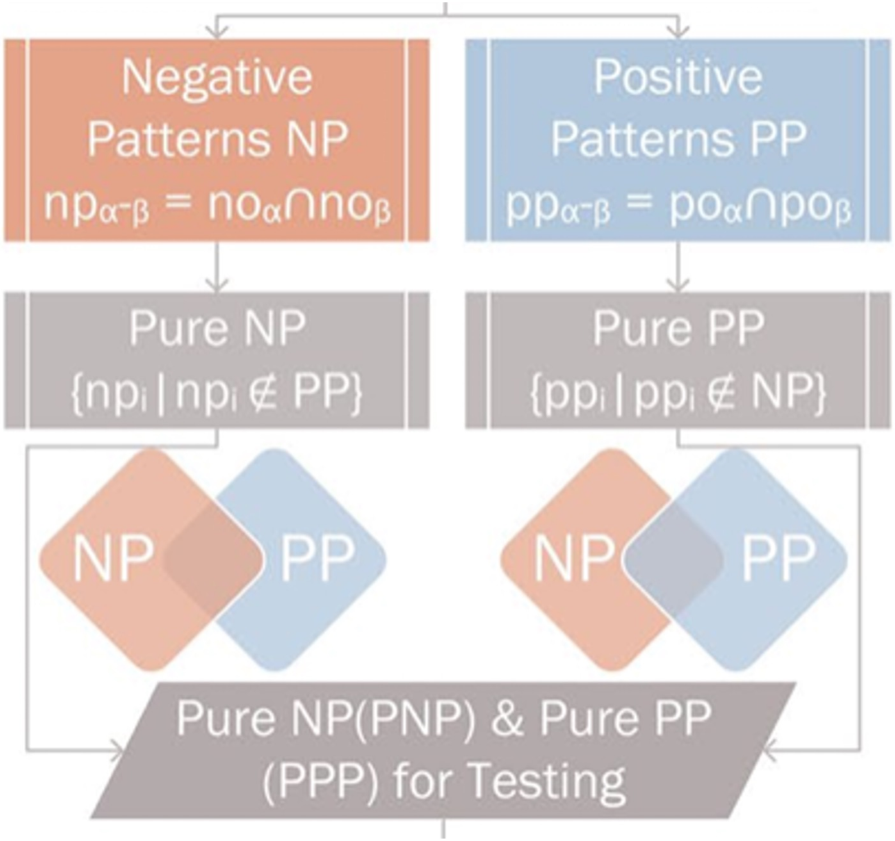
Negative Observations ($\boldsymbol{NO}$): These are instances that belong to the negative class.

Negative Patterns ($\boldsymbol{np_{\alpha-\beta}}$): Similar to positive patterns, these are derived from intersecting different negative observations.

Identifying Distinguishable Patterns

A **Pure Positive Pattern (PPP)** is a positive pattern that does not appear in any negative observation.

A **Pure Negative Pattern (PNP)** is a negative pattern that does not appear in any positive observation.



Identifying Distinguishable Patterns

Ratio of training to testing 5 : 5				
Training				
C0	C1	C2	C3	Class
0	2	4	8	0
0	2	6	7	0
0	3	4	7	0
0	3	5	8	1
1	3	5	8	1
1	2	4	7	0
0	3	5	7	0
Testing				
0	3	4	8	
0	2	6	8	
1	3	4	7	
1	2	5	7	
1	3	4	8	
1	3	5	8	
1	3	5	8	

Negative Patterns (NP)	Pure NP (PNP)
{0,2,4,8},{0,2,6,7},{0,3,4,7}, {1,2,4,7},{0,3,5,7},	{0,2,4,8},{0,2,6,7},{0,3,4,7}, {1,2,4,7},{0,3,5,7},
{0,2,4,8}∩{0,2,6,7}={0,2}, {0,2,4,8}∩{0,3,4,7}={0,4}, {0,2,4,8}∩{1,2,4,7}={2,4}, {0,2,4,8}∩{0,3,5,7}={0}, {0,2,6,7}∩{0,3,4,7}={0,7}, {0,2,6,7}∩{1,2,4,7}={2,7}, {0,2,6,7}∩{0,3,5,7}={0,7}, {0,3,4,7}∩{1,2,4,7}={4,7}, {0,3,4,7}∩{0,3,5,7}={0,3,7}, {1,2,4,7}∩{0,3,5,7}={7}.	{0,2}∉{0,3,5,8},{0,2}∉{1,3,5,8}, {0,4}∉{0,3,5,8},{0,4}∉{1,3,5,8}, {2,4}∉{0,3,5,8},{2,4}∉{1,3,5,8}, ~~{0}∈{0,3,5,8}~~, {0,7}∉{0,3,5,8},{0,7}∉{1,3,5,8}, {2,7}∉{0,3,5,8},{2,7}∉{1,3,5,8}, {4,7}∉{0,3,5,8},{4,7}∉{1,3,5,8}, {0,3,7}∉{0,3,5,8},{0,3,7}∉{1,3,5,8}, {7}∉{0,3,5,8},{7}∉{1,3,5,8}.

Positive Patterns (PP)	Pure PP (PPP)
{0,3,5,8},{1,3,5,8},	{0,3,5,8},{1,3,5,8},
{0,3,5,8}∩{1,3,5,8}={3,5,8}.	{3,5,8}∉{0,2,4,8},{3,5,8}∉{0,2,6,7}, {3,5,8}∉{0,3,4,7},{3,5,8}∉{1,2,4,7}, {3,5,8}∉{0,3,5,7}.

Establishing the Causes

Each observation O_t is given 3 scores:

- **Positive score(PS_t):** If O_t contains a Pure Positive Pattern (PPP) then
$$PS_t = PS_t + |PO^{ppp}| \cdot |ppp|$$
$|ppp|$ - number of features in PPP,
PO^{ppp} - number of Positive observations containing PPP

- **Negative score(NS_t):** If O_t contains a Pure Negative Pattern (PNP) then
$$NS_t = NS_t + |NO^{pnp}| \cdot |pnp|$$
$|ppp|$ - number of feature in PNP,
NO^{pnp} - number of Positive observations containing PNP

- **Novelty score(NT_t):** This represents a neutral count when no known positive or negative patterns are found.
$$NT_t = |O_{TR}|$$
O_{TR} - number of training observations

Understanding Results of Analytics

The authors evaluate the performance of the TC method by three metrics:

- **Precision.** Measures the accuracy of positive predictions.
- **Recall.** Measures the ability to find all relevant instances (true positives).
- **Area Under Curve (AUC).** Assesses the model's ability to discriminate between positive and negative classes.

AUC value	Discrimination
< 0.5	no discrimination
0.5 to 0.7	acceptable
0.7 to 0.9	excellent
> 0.9	outstanding

Understanding Results of Analytics

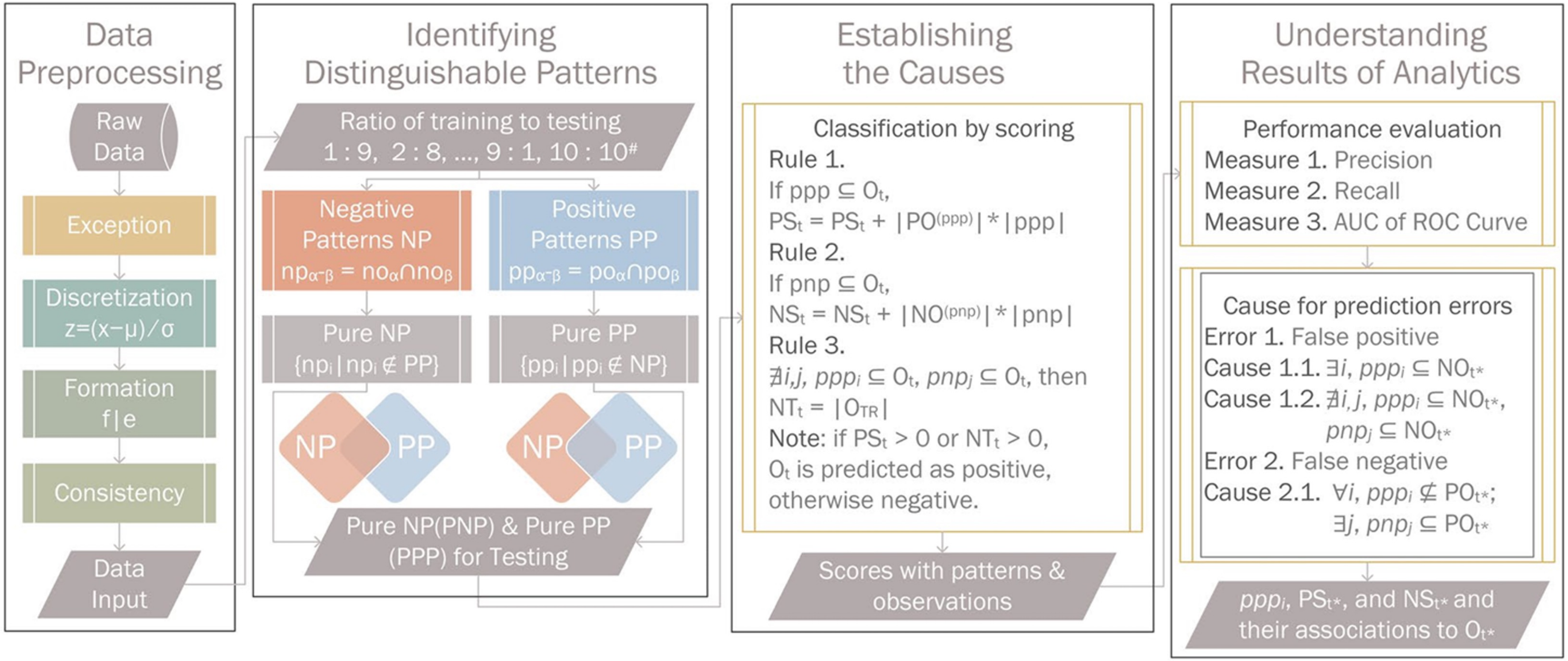
False positive error is caused by:

- NO_t incorrectly containing pure positive patterns;
- NO_t being novel, i.e. lacking known Pure Positive Patterns and Pure Negative Patterns.

False negative error is caused by:

- PO_t containing Pure Negative Patterns, yet lacking Pure Positive Patterns.

TC method step-by-step



Experiments & Results

Performance on BCWO dataset

RA	RE	PR	AU
1:9	1.000	0.701	0.905
2:8	0.994	0.777	0.976
3:7	0.993	0.760	0.988
4:6	1.000	0.777	0.996
5:5	1.000	0.752	0.997
6:4	1.000	0.805	0.997
7:3	1.000	0.810	0.996
8:2	1.000	0.833	0.998
9:1	1.000	0.765	0.999
10:10[#]	1.000	1.000	1.000

No. features: 9 + 1 target class label

No. instances: 699

RA: Ratios of training to testing; **RE**: Recall; **PR**: Precision; **AU**: AUC of ROC Curve.

Performance on CMC dataset

RA	RE	PR	AU
1:9	1.000	0.362	0.432
2:8	1.000	0.282	0.428
3:7	1.000	0.203	0.566
4:6	0.944	0.259	0.662
5:5	0.843	0.321	0.653
6:4	0.763	0.442	0.676
7:3	0.703	0.517	0.673
8:2	0.844	0.163	0.727
9:1	0.000	0.000	0.500
10:10[#]	1.000	1.000	1.000

No. features: 9 + 1 target class label

No. instances: 1473

RA: Ratios of training to testing; **RE**: Recall; **PR**: Precision; **AU**: AUC of ROC Curve.

Results

- The TC method performed well on the Breast Cancer Wisconsin (BCWO) dataset, handling faulty labels and imbalanced data **effectively**.
- On the Contraceptive Method Choice (CMC) dataset, results **improved** with larger training to testing ratios, but the inherent inconsistency in the data remained a **challenge**.

Conclusion

The Transparent Classification method shows promise in both machine learning and data analysis, particularly in **uncovering patterns** in complex datasets and **detecting labeling errors**.

References

- Pai, HT., Hsu, CC. Explainable analytics: understanding causes, correcting errors, and achieving increasingly perfect accuracy from the nature of distinguishable patterns. *Sci Rep* 12, 18368 (2022). https://doi.org/10.1038/s41598-022-19650-2
- Gabriela Alexe, Peter L. Hammer. Spanned patterns for the logical analysis of data. *Discrete Applied Mathematics*, Volume 154, Issue 7, 2006, pp. 1039-1049, ISSN 0166-218X, https://doi.org/10.1016/j.dam.2005.03.031

Thanks!