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

Hao-Ting Pai (✉ greatpai@hotmail.com)

National Yunlin University of Science and Technology <https://orcid.org/0000-0002-6993-1343>

Chung-Chian Hsu

National Yunlin University of Science and Technology

Article

Keywords:

Posted Date: December 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1130288/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

In addition to pursue accurate analytics, it is invaluable to clarify how and why inaccuracy exists. We propose a transparent classification method (TC). In training, we discover patterns from positive and negative observations respectively; next, patterns are excluded if they appear in both types. In testing, observations are scored by the pure patterns and connected like social networks. Based on set theory, pure patterns have explanatory power for distinguishing tangled relationship between negative and positive observations. Experimental results demonstrate that TC can identify all positive (e.g., malignant) observations at low ratios of training to testing, e.g., 1:9 in Breast Cancer Wisconsin (Original) and 3:7 in Contraceptive Method Choice dataset. Without fine-tuned parameters and random selection, TC eliminates uncertainty of the methodology. TC can visualize causes, and therefore, prediction errors are traceable and can be corrected. Further, TC shows potential of identifying whether the ground truth is incorrect (e.g., diagnostic errors).

Main Text

Accurate prediction plays a pivotal role in analytics; however, in reality people usually face the challenge of explaining how and why the prediction is inaccurate¹⁻². By the survey³, outpatient diagnostic errors are at a rate of 5.08% (around 12 million US adults) per year. Even reducing only 1% errors, the lives of million people are saved. We consider three major issues of errors. First, *fault in data*, human mistakes or defective machines can produce faulty data. Without domain knowledge, such a fault is difficult to correct. Nevertheless, we should remove inconsistency, i.e., observations of positive class are identical to those of negative one. In addition, positive and negative observations may have similar patterns that are inextricably interwoven, e.g., people with similar profiles may exhibit different behavior. Lim *et al.*⁴ show the contraceptive method choice (CMC) dataset⁵ is the most difficult to classify, and especially the minimum error rates are greater than 0.4. Second, *mismatches between data and methods*, data contains categorical (e.g., country), numerical (e.g., age), or both types, which have natural constraints on analysis. For categorical values, only the number of items and the mode are statistically relevant⁶ so that a numerical-orientated method is inherently inadequate. For numerical values, they can be transformed into categorical ones by discretization⁷, which have been widely applied to knowledge discovery and data mining (KDD) applications⁸. However, bias occurs if categories are unrepresentative of numerical values. Third, *big data challenge*, the complexity of data is determined by the number of rows and features (columns). Particularly, computation tasks increase rapidly with the number of features, as known curse of dimensionality (CoD)⁹. To cope with CoD, dimension reduction and feature selection methods reduce the complexity by extracting information that is practical for classification and cluster analysis. The extraction, a trade-off between efficiency and effectiveness, may involve pruning large amounts of data. Actually, such a pruning manner may have pitfalls¹⁰ or miss clues to errors.

Results

We conducted experiments with two datasets: Breast Cancer Wisconsin (Original) (BCWO) and Contraceptive Method Choice (CMC) ⁵. Figure 2 (A)(B) and (C)(D) show the results of BCWO and CMC respectively. In Figure 2 (A), TC achieves perfect recall (i.e., recall=1.0) at the lowest ratio (i.e., 1:9) and 7 other ratios. It means TC is not only accurate in small amounts of data but also stable when data increases. TC has one error at the ratios 2:8 and 3:7 because the positive observation PO_{223} is predicted as a negative one. Since the ratio 4:6, PO_{223} belongs to training data and is not used in testing. For further exploration, at the ratio 10:10[#], PO_{223} is irrelevant to PPP except itself. Indeed, PO_{223} is related to PP that are also relevant to NO. PPP can eliminate impingement of PO_{223} on other observations. Novel observations are regarded as positive ones if they are unrelated to both PPP and PNP. Some of NO are predicted as positive so that TC gets lower performance in terms of PR. With increased training observations, TC may obtain sufficient PP or NP to provide accurate results and their causes, e.g., at the ratio 1:9, PO_{627} has a novelty score of 70 by rule 3; at the ratio 2:8, PO_{627} contains 3 patterns in PPP, namely, ['8|0.8'], ['6|1.5'], '8|0.8'], and ['1|0.9', '8|0.8'] and has a positive score of 16 by rule 1. Figure 2 (B) shows the suitable granularity of discretization gives broad vision for discovering more clues. At the ratio 1:9 TC has one error, which is also caused by PO_{223} . At the ratio 2:8, PO_{223} has 1 pattern ['3|0.0', '5|0.0'] in PPP and is related to PO_{33} and PO_{102} . At the ratio 3:7, PO_{223} has 1 pattern ['3|0.0', '5|0.0'] in PPP and related to PO_{33} , PO_{102} , and PO_{143} . Since the ratio 4:6, PO_{223} belongs to training data and is not used in testing. In Figure 2 (C) (D) at the ratios 1:9 and 2:8, RE and PR are zero due to none of negative observations in training (TRN). At the ratio 3:7, insufficient TRN leads to extremes as perfect RE but relatively low PR. The error rate is less than 0.4 at the ratio 7:3. In CMC, profile data is incapable of explaining behavior because (D) indicates 33% (i.e., (1473-980)/1473) cases are inconsistent inherently. In Figure 3, we provide visualization of TC. Compared to images, analysis of categorical and numerical data was difficult to visualize how causes are related to results.

Discussion

Figure 3 (A) shows TC successfully predicts O_{104} is positive because O_{104} is related to the six pure positive patterns that are obtained from respective positive observations of training. Moreover, the thickness of lines represents the score as well as the positive degree. Figure 3 (B) presents the association between O_{104} and others. In the O_{104} -headed group, the most common positive pattern is '0|1.3', and O_{24} and O_{57} have significant influence on O_{104} . Figure 4 (A) (B) (C) illustrate TC is tolerant of faulty class labels (ground truth) in terms of testing, training, and both cases. Figure 4 (D) demonstrates TC can correct errors in class labels. Specifically, observations of faulty class labels have extreme values of PS or NS, which becomes significant with increasing data. Based on profound knowledge of most cases, TC would be useful for checking whether the original judgement (e.g., diagnosis) deserves further inspection.

Methods

We propose a method of transparent classification, named TC, which not only pursues accuracy but also clarifies cause of inaccuracy. Further, the design principles of TC ensure reproducibility¹¹. Figure 1 shows processes of TC. In *data preprocessing*, TC handles missing values and mixed values. Without involving randomness and reduction, TC delivers intrinsic nature of data. In *identifying distinguishable patterns*, TC finds patterns from training observations, which are used for predicting which class a test observation belongs to, e.g., a malignant or benign tumor. Through *increasing ratios of training to testing*, TC represents the forest and the trees, and input data is given in sequence. To avoid CoD, TC finds patterns by intersecting pairwise observations in each of the classes, which possess essential features of data in miniature. In the worst case of TC, n observations produce only patterns. By contrast, KDD faces the challenge of CoD, i.e., given the lowest threshold, k items yield 2^k itemsets¹², and large amounts of itemsets are pruned if the threshold is high. In *positive patterns* (PP), TC obtains PP from positive training observations (PO). For *pure PP* (PPP), TC excludes any positive pattern that also appears in negative training observations (NO). By set theory, the exclusion implies none of PPP is included in any of NO and hence TC can distinguish between PO and NO. Analogously, *Negative patterns* (NP) and *pure NP* (PNP) are the counterpart of PP and PPP. Without involving either fine-tuned parameters or random selection, TC eliminates uncertainty of the methodology. In *establishing the causes*, TC accumulates positive, negative, and novel degrees of a test observation O_t by rule 1, 2, and 3 which associate patterns with the observation and provide obvious clues for judgement. In rule 1, O_t containing patterns in PPP gets a positive score (PS). In rule 2, O_t containing patterns in PNP gets a negative score (NS). In rule 3, O_t containing none of patterns in PPP and PNP is considered novel and gets a novelty score (NT) equal to the number of training observations. In *understanding results of analytics*, we evaluate performance of TC by three measures, *Precision*, *Recall*, and *AUC (Area under Curve)*^{8,12}. According to the standard of diagnostic medicine¹³: $AUC=0.5$, no discrimination; $0.7 \leq AUC < 0.8$, acceptable; $0.8 \leq AUC < 0.9$, excellent; and $0.9 \leq AUC \leq 1$, outstanding. In *cause for prediction errors*, error 1 (false positive¹⁴) occurs if O_t is predicted as positive but actually negative, denoted by NO_{t*} . By cause 1.1, NO_{t*} contains pure positive patterns although it should not. By cause 1.2, NO_{t*} is novel, namely, containing no pattern in PPP and PNP. Error 2 (false negative¹⁴) occurs if O_t is predicted as negative but actually positive, denoted by PO_{t*} . By cause 2.1, PO_{t*} contains pure negative patterns but no pure positive pattern although it should. Prediction errors occur due to insufficient training data or labelling errors in training data. Increasing training data helps to reduce prediction errors. If the portion of labelling errors is small, TC has the potential of identifying labelling errors. Specifically, false negatives usually have a small NS.

Declarations

Acknowledgments: We thank the editor and reviewers for offering comments on the manuscript.

Funding: This work was supported by the Ministry of Science and Technology (MOST) in Taiwan.

MOST (grant no. 110-2410-H-224-037-): H.T.P.

MOST (grant no. 110-2410-H-224-009): C.C.H.

Author contributions:

Conceptualization: H.T.P. conceived the idea and C.C.H. provided for guidance.

Methodology: H.T.P.

Experiments: H.T.P.

Visualization: H.T.P.

Funding acquisition: H.T.P., C.C.H.

Writing – original draft: H.T.P.

Writing – review & editing: H.T.P., C.C.H.

Competing interests: Authors declare that they have no competing interests.

References

1. Description, prediction, explanation. *Nat Hum Behav* **5**, 1261 (2021).
2. D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, G.-Z. Yang, XAI—Explainable artificial intelligence. *Sci. Robot.* **4**, eaay7120 (2019).
3. H. Singh, A. N. Meyer, E. J. Thomas, The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ Qual. Saf.* **23**, 727-731 (2014).
4. T. S. Lim, W. Y. Loh, Y. S. Shih, A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Mach. Learn.* **40**, 203-228 (2000).
5. D. Dheeru, C. Graff, UCI Machine Learning Repository (2019); <http://archive.ics.uci.edu/ml>.
6. S. S. Stevens, On the theory of scales of measurement. *Science*. **103**, 677-680 (1946).
7. S. Garcia, J. Luengo, J. A. Sáez, V. Lopez, F. Herrera, F. A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning. *IEEE Trans. Knowl. Data Eng.* **25**, 734-750 (2012).
8. P. N. Tan, M. Steinbach, V. Kumar, *Introduction to Data Mining* (Pearson, 2020).
9. N. Altman, M. Krzywinski, The curse(s) of dimensionality. *Nat Methods*. **15**, 399-400 (2018).

10. A. Jain, D. Zongker, Feature selection: Evaluation, application, and small sample performance. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 153-158 (1997).
11. M. McNutt, Reproducibility. *Science*. **343**, 229 (2014).
12. M. J. Zaki, W. Meira Jr, *Data Mining and Machine Learning: Fundamental Concepts and Algorithms* (Cambridge University Press, 2020).
13. J. N. Mandrekar, Receiver operating characteristic curve in diagnostic test assessment. *J. Thorac. Oncol.* **5**, 1315-1316 (2010).
14. National Cancer Institute (2021); <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/false-positive-test-result>.

Figures

Figure 1

Processes of the transparent classification. (A) Exception treats missing values as categorical values instead of guesses. Discretization transfers numerical values to categorical ones by using z-score, where x are numerical values of a feature, μ is the mean and σ is the standard deviation. Formation defines the relations between features f and categorical values e . Consistency removes the contradictory observations that have identical features but different class labels. (B) Ratio of training to testing divides data into two parts: training and testing. In training, observations are split positive observations PO and negative observations NO . A positive pattern $pp_{\alpha-\beta}$ was discovered by intersecting po_{α} and po_{β} , where $\alpha=1, 2, \dots, n$ and $\beta=\alpha, \alpha+1, \dots, n$, e.g., given $n=3$, then $pp1-1=po1 \cap po1=\{f1|e1,1, f2|e2,2, f3|e3,1\}$, $pp1-2=po1 \cap po2=\{f1|e1,1, f2|e2,2, f3|e3,1\} \cap \{f1|e1,1, f2|e2,1, f3|e3,1\}=\{f1|e1,1, f3|e3,1\}$, $pp1-3=po1 \cap po3$, $pp2-2=po2$, $pp2-3=po2 \cap po3$, and $pp3-3=po3$. Note positive observations themselves are positive patterns, e.g., $pp1-1$. A negative pattern $np_{\alpha-\beta}$ was found by $no_{\alpha} \cap no_{\beta}$, and negative observations themselves are

negative patterns. Pure PP (PPP) excludes $pp\alpha\beta$ that appears in any negative observation. Pure NP (PNP) excludes $np\alpha\beta$ that appears in any positive observation. (C) In testing of an observation O_t , classification by scoring produces five outputs: PS, NS, NT, PSP, and PSO. PS stores observation's positive score by Rule 1: O_t contains a pure positive pattern ppp , increase PS by the number of features in ppp , multiplied by the number of positive observations containing ppp . Rule 2: if O_t contains a pure negative pattern pn , increase NS by the number of features in pn multiplied by the number of negative observations containing pn . Rule 3: if O_t does not contain any ppp and pn , assign NT to the number of training observations. PSP stores $ppp\alpha\beta$ related to O_t . PSO stores the training observations which contain $ppp\alpha\beta$. If PS or NT is greater than 0, classify O_t as positive otherwise negative. (D) Performance evaluation demonstrates the accuracy of TC. Cause for prediction errors, based on set theory, provides rational explanations for errors caused by TC.

A

RA	RE	PR	AU	ER	TP	TN	FP	FN	TRP	TRN	TEP	TEN	UPP	PPP
1:9	1.000	0.701	0.905	0.140	206	335	88	0	35	35	206	423	251	201
2:8	0.994	0.777	0.976	0.093	178	329	51	1	62	78	179	380	592	468
3:7	0.993	0.760	0.988	0.100	152	288	48	1	88	122	153	336	1007	845
4:6	1.000	0.777	0.996	0.079	115	271	33	0	126	154	115	304	1630	1364
5:5	1.000	0.752	0.997	0.077	82	240	27	0	159	191	82	267	2266	1880
6:4	1.000	0.805	0.997	0.057	66	198	16	0	175	244	66	214	2599	2163
7:3	1.000	0.810	0.996	0.052	47	152	11	0	194	295	47	163	2968	2493
8:2	1.000	0.833	0.998	0.050	35	98	7	0	206	353	35	105	3225	2710
9:1	1.000	0.765	0.999	0.057	13	53	4	0	228	401	13	57	3723	3162
10:10 [#]	1.000	1.000	1.000	0.000	241	458	0	0	241	458	241	458	4023	3422

RA: Ratios of training to testing; RE: Recall; PR: Precision; AU: AUC of ROC Curve; ER: Error rate; TP: True positive; TN: True negative; FN: False negative; FP: False positive; TRP: Positive observations of training; TRN: Negative observations of training; TEP: Positive observations of testing; TEN: Negative observations of testing; UPP: Unique positive patterns; PPP: Pure positive patterns. In RA, for comprehensive exploration, 10:10[#] means we respectively give the entire data in training and test.

B

RA	RE	PR	AU	ER	TP	TN	FP	FN	TRP	TRN	TEP	TEN	UPP	PPP
1:9	0.995	0.670	0.904	0.162	205	322	101	1	35	35	206	423	359	273
2:8	1.000	0.731	0.975	0.118	179	314	66	0	62	78	179	380	861	677
3:7	1.000	0.750	0.986	0.104	153	285	51	0	88	122	153	336	1390	1153
4:6	1.000	0.737	0.992	0.098	115	263	41	0	126	154	115	304	2308	1889
5:5	1.000	0.766	0.997	0.072	82	242	25	0	159	191	82	267	3211	2564
6:4	1.000	0.815	0.998	0.054	66	199	15	0	175	244	66	214	3666	2959
7:3	1.000	0.839	0.998	0.043	47	154	9	0	194	295	47	163	4162	3371
8:2	1.000	0.875	0.998	0.036	35	100	5	0	206	353	35	105	4555	3697
9:1	1.000	0.765	1.000	0.057	13	53	4	0	228	401	13	57	5286	4350
10:10 [#]	1.000	1.000	1.000	0.000	241	458	0	0	241	458	241	458	5675	4687

C

RA	RE	PR	AU	ER	TP	TN	FP	FN	TRP	TRN	TEP	TEN	UPP	PPP
1:9	1.000	0.362	0.432	0.638	456	0	803	0	140	0	456	803	1539	1539
2:8	1.000	0.282	0.428	0.718	316	0	803	0	280	0	316	803	3555	3555

3:7	1.000	0.203	0.566	0.795	198	3	778	0	398	22	198	781	5539	4252
4:6	0.944	0.259	0.662	0.652	187	105	536	11	398	162	198	641	5539	2762
5:5	0.843	0.321	0.653	0.549	167	148	353	31	398	302	198	501	5539	2141
6:4	0.763	0.442	0.676	0.425	151	171	191	47	398	441	198	362	5539	1798
7:3	0.703	0.517	0.673	0.390	121	135	113	51	424	555	172	248	5938	1783
8:2	0.844	0.163	0.727	0.514	27	109	139	5	564	555	32	248	8345	2693
9:1	0.000	0.000	0.500	0.536	0	65	75	0	596	663	0	140	8744	2728
10:10 [#]	1.000	1.000	1.000	0.000	596	803	0	0	596	803	596	803	8744	2566

D

RA	RE	PR	AU	ER	TP	TN	FP	FN	TRP	TRN	TEP	TEN	UPP	PPP
1:9	1.000	0.367	0.503	0.633	324	0	558	0	98	0	324	558	1480	1480
2:8	1.000	0.288	0.515	0.712	226	0	558	0	196	0	226	558	3524	3524
3:7	1.000	0.213	0.632	0.770	143	15	528	0	279	15	143	543	5580	4293
4:6	0.986	0.283	0.712	0.611	141	88	357	2	279	113	143	445	5580	2723
5:5	0.958	0.369	0.751	0.490	137	113	234	6	279	211	143	347	5580	1978
6:4	0.888	0.498	0.779	0.367	127	121	128	16	279	309	143	249	5580	1564
7:3	0.866	0.569	0.778	0.320	103	97	78	16	303	383	119	175	5994	1572
8:2	0.857	0.165	0.798	0.480	18	84	91	3	401	383	21	175	8192	2355
9:1	0.000	0.000	0.500	0.592	0	40	58	0	422	460	0	98	8560	2320
10:10 [#]	1.000	1.000	1.000	0.000	422	558	0	0	422	558	422	558	8560	2120

Figure 2

Performance of TC in distinguishing between observations. (A) (B) In Breast Cancer Wisconsin (Original) data set (BCWO), we map class values “malignant” to “1” and “benign” to “0”. In case (A), the granularity of discretization is to the first decimal place, e.g. $1.68 \approx 1.6$, while in case (B) we take an integer for the granularity, e.g., $1.68 \approx 1$. (C) (D) In Contraceptive Method Choice data set (CMC), we map class values “1=No-use” to “1”, “2=Long-term” to “0”, and “3=Short-term” to “0”. For (C) and (D), we set the same

granularity as that of (A) and (B), respectively. For consistency, we remove observations that have identical features but different class labels. The number of observations is thus reduced from 1473 to 1399 in (C) and from 1473 to 980 in (D).

Figure 3

Association between patterns and observations in BCWO (A) at the ratio 1:9.

A														
RA	RE	PR	AU	ER	TP	TN	FP	FN	TRP	TRN	TEP	TEN	UPP	PPP
1:9	0.995	0.660	0.904	0.167	202	322	104	1	35	35	203	426	359	273
2:8	1.000	0.718	0.975	0.123	176	314	69	0	62	78	176	383	861	677
3:7	1.000	0.735	0.985	0.110	150	285	54	0	88	122	150	339	1390	1153
4:6	1.000	0.718	0.991	0.105	112	263	44	0	126	154	112	307	2308	1889
5:5	1.000	0.738	0.996	0.080	79	242	28	0	159	191	79	270	3211	2564
6:4	1.000	0.778	0.996	0.064	63	199	18	0	175	244	63	217	3666	2959
7:3	1.000	0.786	0.996	0.057	44	154	12	0	194	295	44	166	4162	3371
8:2	1.000	0.800	0.995	0.057	32	100	8	0	206	353	32	108	4555	3697
9:1	1.000	0.588	0.985	0.100	10	53	7	0	228	401	10	60	5286	4350
10:10 [#]	1.000	1.000	1.000	0.000	238	461	0	0	238	461	238	461	5600	4525
B														
RA	RE	PR	AU	ER	TP	TN	FP	FN	TRP	TRN	TEP	TEN	UPP	PPP
1:9	0.981	0.699	0.908	0.145	202	336	87	4	32	38	206	423	318	219
2:8	1.000	0.762	0.980	0.100	179	324	56	0	59	81	179	380	806	590
3:7	1.000	0.789	0.989	0.084	153	295	41	0	85	125	153	336	1318	1040
4:6	1.000	0.777	0.994	0.079	115	271	33	0	123	157	115	304	2226	1760
5:5	1.000	0.781	0.997	0.066	82	244	23	0	156	194	82	267	3120	2416
6:4	1.000	0.835	0.998	0.046	66	201	13	0	172	247	66	214	3567	2798
7:3	1.000	0.839	0.997	0.043	47	154	9	0	191	298	47	163	4060	3202
8:2	1.000	0.875	0.998	0.036	35	100	5	0	203	356	35	105	4451	3522
9:1	1.000	0.765	1.000	0.057	13	53	4	0	225	404	13	57	5190	4153
10:10 [#]	1.000	1.000	1.000	0.000	238	461	0	0	238	461	238	461	5578	4485
C														
RA	RE	PR	AU	ER	TP	TN	FP	FN	TRP	TRN	TEP	TEN	UPP	PPP
1:9	0.980	0.689	0.907	0.149	199	336	90	4	32	38	203	426	318	219
2:8	1.000	0.749	0.980	0.106	176	324	59	0	59	81	176	383	806	590
3:7	1.000	0.773	0.988	0.090	150	295	44	0	85	125	150	339	1318	1040
4:6	1.000	0.757	0.992	0.086	112	271	36	0	123	157	112	307	2226	1760
5:5	1.000	0.752	0.996	0.074	79	244	26	0	156	194	79	270	3120	2416
6:4	1.000	0.797	0.996	0.057	63	201	16	0	172	247	63	217	3567	2798
7:3	1.000	0.786	0.994	0.057	44	154	12	0	191	298	44	166	4060	3202
8:2	1.000	0.800	0.993	0.057	32	100	8	0	203	356	32	108	4451	3522
9:1	1.000	0.588	0.985	0.100	10	53	7	0	225	404	10	60	5190	4153
10:10 [#]	1.000	1.000	1.000	0.000	235	464	0	0	235	464	235	464	5504	4331
D														
Faults in testing (cases 697, 698, and 699)					Faults in training (cases 6, 13, and 15)					Faults in training and testing (cases 6, 13, 15, 697, 698, and 699)				
RA	O	NT	PS	NS	RA	O	NT	PS	NS	RA	O	NT	PS	NS
1:9	697	0	42	0	1:9	223	0	0	50	1:9	697	0	33	0
1:9	698	0	9	0	1:9	286	0	0	13	1:9	698	0	9	0
1:9	699	0	23	0	1:9	345	0	0	4	1:9	699	0	10	0
...	...	...	...	...	1:9	569	0	0	4	...	...	...	...	...
9:1	697	0	2672	0	2:8	223	0	4	875	9:1	697	0	2293	0
9:1	698	0	134	0	2:8	286	0	123	9	9:1	698	0	134	0
9:1	699	0	1016	9	2:8	345	0	12	0	9:1	699	0	1016	9
10:10 [#]	695	0	0	95204	2:8	569	0	13	0	10:10 [#]	695	0	0	95204
10:10 [#]	696	0	0	3537383	10:10 [#]	5	0	0	143716	10:10 [#]	696	0	0	3537383
10:10 [#]	697	0	0	99	10:10 [#]	6	0	0	81	10:10 [#]	697	0	0	117
10:10 [#]	698	0	0	106	10:10 [#]	13	0	0	726	10:10 [#]	698	0	0	106
10:10 [#]	699	0	0	140	10:10 [#]	15	0	0	81	10:10 [#]	699	0	0	140

Figure 4

Tolerance to faulty class labels in BCWO (B). (A) For the case of testing, we change class labels of observations (i.e., 697, 698, and 699) from “1” to “0”. (B) For the case of training, class labels of observations (i.e., 6, 13, and 15) are changed from “1” to “0”. (C) For both cases, we change class labels of observations (i.e., 6, 13, 15, 697, 698, and 699) from “1” to “0”. (D) Although faults in testing, TC can still classify O697, O698, and O699 as PO since the ratio 1:9. Regarding faults in training, the change

results in an additional three errors at the ratio 1:9, i.e., O286, O345, and O569. Since the ratio 2:8, the three errors are eliminated due to increased training data. Although faults in testing and training, TC can still classify O697, O698, and O699 as PO since the ratio 1:9. Note at the ratio 10:10#, O697, O698, and O699 belong to training data so that their PS are zero; nevertheless, we can identify them by NS.