Also „Research conducted as part of the PREFERML project (project partner SICK AG). Funded by the German Federal Ministry of Education and Research, funding line “Forschung an Fachhochschulen mit Unternehmen (FHProfUnt)“, contract number 13FH249PX6

EVALUATION OF FILTER METHODS FOR FEATURE SELECTION BY USING REAL MANUFACTURING DATA

Authors: Alexander Gerling, Holger Ziekow, Ulf Schreier, Christian Seiffer, Andreas Hess and Djaffar Ould Abdeslam

Presenter: Alexander Gerling, Hochschule Furtwangen

Email: alexander.gerling@hs-furtwangen.de

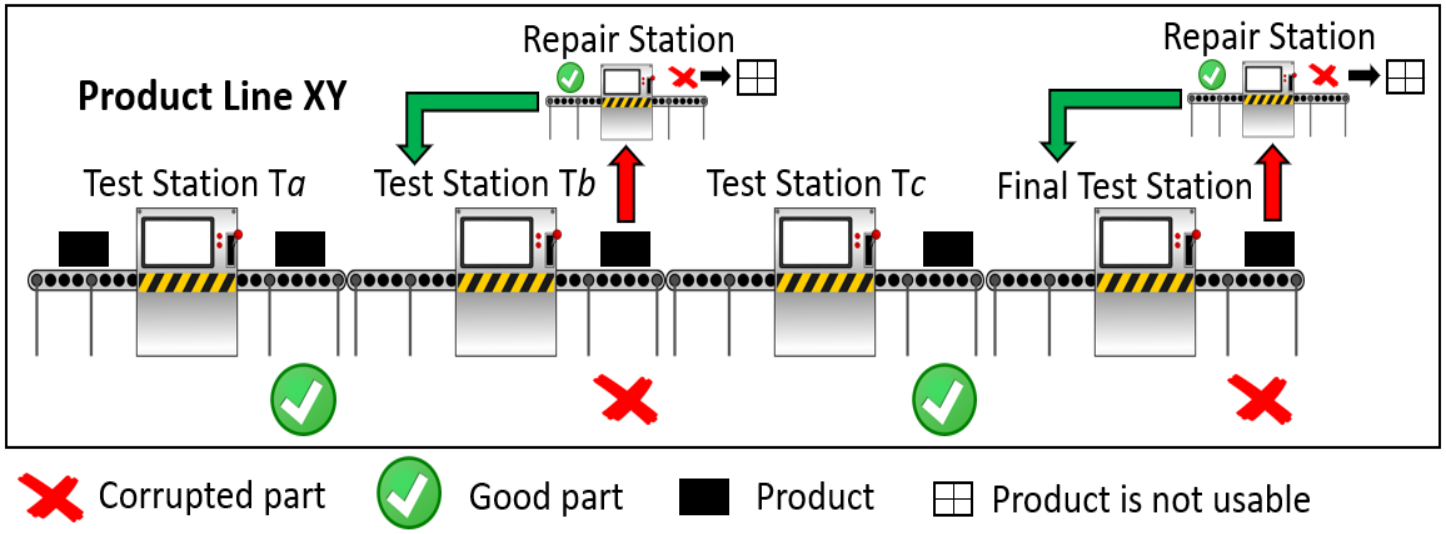
Person

- 2018, Master Degree at Furtwangen University
- Working at a research project „PREFERML “ at Furtwangen University since November 2018
- Ph.D. Student at the Université de Haute-Alsace since March 2019

Agenda

- Manufacturing Domain
- Motivation
- Filter Selection Method
- Metrics and Pseudo Code
- Datasets and Results
- Feature Selection Approach A – C
- Execution time
- Result overview
- Conclusion

Manufacturing Domain



col1	col2	col3	col4	col5	col6	col7	col8	col9	col10
1.3	60	82	48	15.8	35	927	470	83	2
0.6	10	40	34	7.8	20	985	479	82	1
95	59	90	39	17.5	24	152	473	28	1
87	63	35	45	16.2	36	842	409	25	2
87	71	80	49	9.7	33	962	493	28	1
39.52	10	81	36	13.75	33	985	191	26	3
97	52	93	44	11.55	39	260	483	86	2
88	11	67	43	8.3	20	228	209	78	2
1.2	11	28	47	17.1	27	655	217	25	4
99	13	78	45	18.5	29	273	377	84	4
38.12	14	53	42	7	29	519	238	28	4
82	12	36	37	18.5	31	545	496	33	1
82	11	12	44	7.4	25	146	214	38	4
0.6	59	10	37	7.5	34	796	324	80	2
1.2	53	36	40	7	26	111	305	27	3
0.4	13	18	40	8.9	37	673	232	29	1
99	11	95	44	19.9	37	643	470	26	3
92	14	41	36	8.3	27	701	486	28	4
0.8	66	63	35	11.65	31	764	246	78	4
1	56	42	45	18.5	22	929	441	31	4
43.92	69	42	42	18.1	22	565	303	26	4
1.1	10	21	42	9.1	24	561	165	78	4
0.6	14	82	45	18.8	20	800	455	81	5
99	14	73	36	8.7	32	107	304	30	3
96	14	84	30	19.1	35	144	462	35	4
43.12	11	67	36	9.1	37	340	481	88	3
95	61	67	49	11.75	30	631	470	36	2
85	54	14	47	16.6	31	523	203	32	4
44.62	56	73	47	17.2	39	381	125	82	1
1.2	65	22	38	7.8	24	113	165	37	3
44.22	10	34	42	9.1	26	328	110	36	5
0.6	13	78	47	12.45	33	934	425	83	1
40.72	61	93	36	12.75	34	509	200	29	2
86	64	21	34	11.55	38	426	450	26	1

- A single Test Station has often many features to test
- A Chain with multiple Test Station increases the number of features highly
- A Quality Engineer has to find relevant features of an error

Motivation

Origin Dataset

col1	col2	col3	col4	col5	col6	col7	col8	col9	col10
1.3	60	82	48	15.8	35	927	470	83	2
0.6	10	40	34	7.8	20	985	479	82	1
95	59	90	39	17.5	24	152	473	28	1
87	63	35	45	16.2	36	842	409	25	2
87	71	80	49	9.7	33	962	493	28	1
39.52	10	81	36	13.75	33	985	191	26	3
97	52	93	44	11.55	39	260	483	86	2
88	11	67	43	8.3	20	228	209	78	2
1.2	11	28	47	17.1	27	655	217	25	4
99	13	78	45	18.5	29	273	377	84	4
38.12	14	53	42	7	29	519	238	28	4
82	12	36	37	18.5	31	545	496	33	1
82	11	12	44	7.4	25	146	214	38	4
0.6	59	10	37	7.5	34	796	324	80	2
1.2	53	36	40	7	26	111	305	27	3
0.4	13	18	40	8.9	37	673	232	29	1
99	11	95	44	19.9	37	643	470	26	3
92	14	41	36	8.3	27	701	486	28	4
0.8	66	63	35	11.65	31	764	246	78	4
1	56	42	45	18.5	22	929	441	31	4
43.92	69	42	42	18.1	22	565	303	26	4
1.1	10	21	42	9.1	24	561	165	78	4
0.6	14	82	45	18.8	20	800	455	81	5
99	14	73	36	8.7	32	107	304	30	3
96	14	84	30	19.1	35	144	462	35	4

Most important Feature

col2	col5	col8
60	15.8	470
10	7.8	479
59	17.5	473
63	16.2	409
71	9.7	493
10	13.75	191
52	11.55	483
11	8.3	209
11	17.1	217
13	18.5	377
14	7	238
12	18.5	496
11	7.4	214
59	7.5	324
53	7	305
13	8.9	232
11	19.9	470
14	8.3	486
66	11.65	246
56	18.5	441
69	18.1	303
10	9.1	165
14	18.8	455
14	8.7	304
14	19.1	462

- High number of features has to be analyzed manually
- Manual filtering could ignore important feature
- Better or same model performance
- Improve interpretability of product error

Filter Selection Method

Used Methods:

- **ANOVA**
- **Kendall's rank coefficient**
- **Permutation Feature Importance**

- Advantage of better time performance in comparison to wrapper methods
- Classifier independent, therefore more flexibility to choose a different classifier regarding black box optimization.
- Advantage of the filter-based methods is the ability to scale up to high-dimensional datasets

Metrics and Pseudo Code

STANDARD

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

ADJUSTED

$$\text{Expected Benefit Rate (EBR)} = \frac{\mathrm{TP} * \alpha - \mathrm{FP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}}$$

- **α = Adjustable cost factor of how much a correctly identified error in relation to an incorrectly identified error will save us**
- **EBR result shows if an ML model is monetary profitable**

TP = Corrupt part predicted as error FP = Good part predicted as error
FN = Corrupt part predicted as good TN = Good part predicted as good

Pseudo Code

```
1.  Fselect(F, m, ≥r, p, T, V)
2.      S ← F, opt ← −∞
3.      Sort(F, ≥r)
4.      For i = 1 to |F|
5.          C ← {fk ∈ F | k ≤ i}
6.          score ← m(C,p,T,V)
7.          If score > opt and lp < α
8.              opt ← score
9.              S ← C
10.     Return (S)
```

Integrated significance tests in line 7

Datasets and Results

- 25 highly unbalanced datasets used for the experiments
- Dataset hat an instance range from 6887 to 194932
- Lowest good/corrupt product ratio with 0.001228
- Datasets had a feature range from 17 to 133

- EBR result > 0 ; ML model is profitable to use
- EBR result = 0 ; Model may found a relation, but the predicted error probabilities are too low for making an economically reasonable prediction
- EBR result < 0 ; ML model could not find a relation to the error origin

Baseline Results:

- 11 out of 25 Results > 0
- 9 out of 25 Results = 0
- 5 out of 25 Results < 0

Feature Selection Approach A

- Only feature selection with standard parameter for the ML model was used
- We did not consider the unbalanced datasets
- Compared to the baseline, we improved the result in 16 out of 75 experiments based on the EBR optimization.
- Eight deteriorations compared to the baseline
- Deteriorations could be due to a concept drift in the data
- ANOVA selection method was the best for approach A

Feature Selection Approach B

- After feature selection, we used hyperparameter tuning for the best set of features
- 21 out of 75 better results and 9 out of 75 worse results based on the EBR optimization compared to baseline
- 11 out of 75 results and 6 out of 75 got worse based on the EBR optimization compared to approach A
- Advantage to adjust the parameter of the algorithm to provide better results
- Kendall's rank provided the best method if we consider the results from the EBR and MCC optimization

Feature Selection Approach C

- Every model was optimized with hyperparameter tuning within feature selection
- Most changes in the number of features and the difference between the optimization metric
- 16 out of 25 best results based on the EBR and MCC optimization with Kendall's rank
- We could reduce in 21 out of 75 cases the number of features and improve the result by optimizing with the EBR metric.
- Compared to approach B we improve 20 results and got worse in 15 cases based on the EBR results

Execution time

- Approach A best time result = Permutation feature importance in 11 out of 25 cases.
- Approach B best time result = Kendall's rank in 13 out of 25 cases.
- Approach C best time result = ANOVA in 15 out of 25 cases.
- Approach A is the fastest approach ; Average time based with ANOVA 00:03:53 (all 25 datasets)
- Approach C is most time consuming ; Average time based with ANOVA 06:52:26 (all 25 datasets)

Result overview

- Number of tests where best result is with EBR (overall) = 70
- Number of tests where best result is with MCC (overall) = 40

- Number of tests where BEST results is with EBR and Features reduced (overall) = 19
- Number of tests where BEST results is with MCC and Features reduced (overall) = 9

- Number of tests where optimizing with EBR is BETTER than baseline (overall) = 56
- Number of tests where optimizing with MCC is BETTER than baseline (overall) = 46

Conclusion

- A total of 10 Experiments with 25 datasets
- We obtain most of the best results with experiment Approach C
- Most of the best results for the experiment approaches were achieved by using the permutation feature importance selection method
- More features of the dataset can be reduced when using the EBR metric compared to the MCC metric
- Kendall's rank selection method could be used in combination with experiment approach B as fastest method regarding the best possible results