A KNOWLEDGE BASE APPROACH TO LAZY CLASSIFICATION WITH PATTERN STRUCTURES

Abdulrahim Ghazal[1] and Sergei O. Kuznetsov[2]

[1] National Research University Higher School of Economics, Pokrovsky boulevard, 11, 109028, Russia, Moscow
agazal@hse.ru

[2] National Research University Higher School of Economics, Pokrovsky boulevard, 11, 109028, Russia, Moscow
skuznetsov@hse.ru

1 Results

Table 1. Lazy Classification Results.

Exp *min_df*	Accuracy	Precision	Recall	F1 (100)	saved effort (100)
1 0.01	88.5	100	97.6	**98.8**	**88.8**
2 0.02	82.3	100	94.9	97.3	82.4
3 0.03	80.9	100	91.2	89.4	81.0
4 0.04	78.7	100	91.3	95.5	78.7
5 0.05	78.9	100	89.7	94.5	78.9

Table 2. Relaxed Lazy Classification Results.

α	Exp , *min_df*	F1 (100)	saved effort (100)
75	1 0.01	98.0	**92.5**
	2 0.02	97.9	89.1
	3 0.03	97.6	87.4
	4 0.04	97.0	86.5
	5 0.05	96.4	85.4
80	1 0.01	98.0	92.2
	2 0.02	97.9	89.1
	3 0.03	97.6	87.4
	4 0.04	97.0	86.5
	5 0.05	96.4	85.4
85	1 0.01	98.3	91.7
	2 0.02	98.1	88.7
	3 0.03	97.8	87.0
	4 0.04	97.2	86.1
	5 0.05	96.6	85.0
90	1 0.01	**98.6**	90.9
	2 0.02	98.1	87.8
	3 0.03	97.7	86.0
	4 0.04	96.7	84.3
	5 0.05	96.2	83.5
95	1 0.01	98.4	89.8
	2 0.02	97.8	86.9
	3 0.03	97.1	84.8
	4 0.04	96.2	83.2
	5 0.05	95.5	82.0

Table 3. Interval Lazy Classification Results.

Exp , *min_df*	F1 (100)	saved effort (100)
1 0.01	**88.0**	**87.7**
2 0.02	78.4	79.8
3 0.03	76.1	70.6
4 0.04	70.4	65.8
5 0.05	66.5	61.3

Table 4. Max Lazy Classification Results.

Exp , *min_df*	F1 (100)	saved effort (100)
1 0.01	**84.3**	**87.8**
2 0.02	76.0	81.1
3 0.03	68.9	72.3
4 0.04	58.8	67.0
5 0.05	59.9	64.5

Table 5. Min Lazy Classification Results.

Exp , *min_df*	F1 (100)	saved effort (100)
1 0.01	**89.7**	**87.6**
2 0.02	69.5	65.3
3 0.03	75.9	74.9
4 0.04	54.0	59.7
5 0.05	65.7	62.2

Table 6. Relaxed Interval Lazy Classification Results.

α	Exp , *min_df*	F1 (100)	saved effort (100)
75	1 0.01	80.8	**94.1**
	2 0.02	89.8	87.1
	3 0.03	94.1	81.2
	4 0.04	91.8	76.7
	5 0.05	88.0	70.6
80	1 0.01	83.0	93.7
	2 0.02	90.0	87.0
	3 0.03	**94.2**	81.1
	4 0.04	91.8	76.5
	5 0.05	88.1	70.5
85	1 0.01	86.4	93.3
	2 0.02	91.9	86.9
	3 0.03	94.1	81.1
	4 0.04	91.6	76.5
	5 0.05	88.1	70.8
90	1 0.01	83.5	92.9
	2 0.02	91.1	86.7
	3 0.03	93.5	81.2
	4 0.04	92.1	76.7
	5 0.05	87.9	70.5
95	1 0.01	84.7	92.6
	2 0.02	91.3	86.6
	3 0.03	93.8	80.8
	4 0.04	91.6	76.4
	5 0.05	87.8	70.6

Table 7. Relaxed Max Lazy Classification Results.

α	Exp , min_df	F1 (100)	saved effort (100)
75	1 0.01	78.2	94.5
	2 0.02	85.9	87.8
	3 0.03	90.0	82.0
	4 0.04	87.3	76.8
	5 0.05	85.0	70.4
80	1 0.01	80.7	94.4
	2 0.02	85.5	87.5
	3 0.03	90.5	81.6
	4 0.04	88.8	76.9
	5 0.05	86.0	70.6
85	1 0.01	78.0	**94.5**
	2 0.02	83.7	87.4
	3 0.03	90.2	81.4
	4 0.04	89.0	76.7
	5 0.05	86.0	70.6
90	1 0.01	79.7	**94.5**
	2 0.02	86.0	87.4
	3 0.03	90.5	81.8
	4 0.04	89.4	76.3
	5 0.05	86.3	70.6
95	1 0.01	81.8	**94.5**
	2 0.02	84.9	87.5
	3 0.03	**91.7**	81.4
	4 0.04	89.0	76.5
	5 0.05	86.0	70.6

Table 8. Relaxed Min Lazy Classification Results.

α	Exp , min_df	F1 (100)	saved effort (100)
75	1 0.01	84.9	**94.2**
	2 0.02	89.1	87.3
	3 0.03	**94.1**	81.4
	4 0.04	91.3	76.7
	5 0.05	87.4	70.6
80	1 0.01	82.9	**94.2**
	2 0.02	90.1	87.4
	3 0.03	93.7	81.4
	4 0.04	91.1	76.7
	5 0.05	87.5	70.2
85	1 0.01	85.8	93.8
	2 0.02	91.7	87.1
	3 0.03	94.0	81.2
	4 0.04	91.7	76.8
	5 0.05	87.4	70.6
90	1 0.01	84.6	93.2
	2 0.02	90.9	87.1
	3 0.03	93.5	81.4
	4 0.04	91.5	76.7
	5 0.05	87.6	70.3
95	1 0.01	84.9	92.8
	2 0.02	90.8	87.1
	3 0.03	93.4	80.8
	4 0.04	91.6	76.7
	5 0.05	87.4	70.6

Table 9. Comparing Lazy Classification to knowledge Base.

Experiment	Highest F1 (100)	Average time (ms)	Highest F1 using KB (100)	Average Time with KB (ms)
FCA-based classification	98.8	10.4	97.0	0.03
Interval Representation	94.2	1188	93.9	0.04
One-sided Interval (Max)	94.2	1268	93.0	0.04
One-sided Interval (Min)	91.7	1340	91.0	0.04

Table 10. Comparing Lazy Classification to knowledge Base.

Experiment	Highest F1 (100)	Average time (ms)	Highest F1 using KB (100)	Average Time with KB (ms)
FCA-based classification	98.8	10.4	97.0	0.03
Interval Representation	94.2	1188	94.2	0.05
One-sided Interval (Max)	94.2	1268	93.2	0.04
One-sided Interval (Min)	91.7	1340	91.4	0.06

Table 11. Results of Other ML Classification Models

Model	Binary attributes - Highest F1 (100)	Average time (ms)	Real attributes - Highest F1 (100)	Average Time (ms)
GaussianNB	95.1	0.011	97.0	0.011
SVM	96.2	0.027	95.6	0.036
Logistic Regression	95.8	0.02	95.6	0.026
Decision Trees	95.1	0.011	95.2	0.008
Random Forest	86.4	0.03	86.7	0.027